



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS AGRONÓMICAS

ESCUELA DE POSTGRADO

**CALIBRACIÓN DE UN ESPECTRÓMETRO PORTÁTIL DE
INFRARROJO CERCANO PARA PREDECIR NIVELES DE pH,
CONTENIDO DE ARCILLA Y CATIONES INTERCAMBIABLES EN
LOS SUELOS**

*Calibration of a portable near-infrared spectrometer to predict levels of pH, clay
content and exchangeable cations in soils*

Tesis para optar al grado de Magíster en Manejo de Suelos y Aguas

ANGELO BENVENUTO ACEVEDO

Ingeniero Agrónomo

SANTIAGO - CHILE

2022

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS AGRONÓMICAS
ESCUELA DE POSTGRADO

**CALIBRACIÓN DE UN ESPECTRÓMETRO PORTÁTIL DE INFRARROJO
CERCANO PARA PREDECIR NIVELES DE pH, CONTENIDO DE ARCILLA Y
CATIONES INTERCAMBIABLES EN LOS SUELOS**

***CALIBRATION OF A PORTABLE NEAR-INFRARED SPECTROMETER TO
PREDICT THE LEVEL OF pH, CLAY CONTENT AND EXCHANGEABLE CATIONS
IN SOILS***

Tesis presentada como parte de los requisitos para optar al Grado de Magíster en Manejo
de Suelos y Aguas

ANGELO DE PAOLI BENVENUTO ACEVEDO

	Calificación
DIRECTOR DE TESIS	
Osvaldo Salazar Guerrero Ingeniero Agrónomo, MS, Ph.D.	7,0
PROFESORES CONSEJEROS	
Manuel Casanova Pinto Ingeniero Agrónomo, MSc.	6,8
Yasna Tapia Fernández Ingeniera en Alimentos, Dr.	6,4

Santiago, Chile
2022

AGRADECIMIENTOS

ÍNDICE DE CONTENIDO

INTRODUCCIÓN	1
Propiedades de suelo y técnicas pedométricas	1
Importancia de la estimación de pH, contenido de arcilla y cationes de intercambio en los suelos.....	3
HIPÓTESIS	4
OBJETIVOS	4
Objetivo general	4
Objetivos específicos	4
MATERIALES Y MÉTODOS	5
Área de estudio.....	5
Toma de muestras	5
Calibración y análisis estadístico	5
Determinación del error por contenido de agua de la muestra (EPO)	7
Calibración de muestras	9
RESULTADOS Y DISCUSIÓN	11
Características de las muestras de suelo.....	11
Tratamiento de espectros.....	13
Calibración de muestras en seco	17
Creación de matriz de corrección de humedad (EPO)	21
CONCLUSIONES	31
LITERATURA CITADA	32

ÍNDICE DE CUADROS

Página

Cuadro 1. Propiedades y estadígrafos de los suelos utilizados en el estudio.....	13
Cuadro 2. Predicción de pH, contenido de arcilla y cationes de intercambio con mínimos cuadrados parciales (PLSR), modelo cubista (CM), usando espectros cortados y transformados.....	19
Cuadro 3. Predicción de pH, contenido de arcilla y cationes de intercambio con mínimos cuadrados parciales (PLSR) y modelo cubista (CM) usando espectros sin transformación.....	21
Cuadro 4. Longitudes de onda usadas en los modelos de PLSR y CM con espectro completo.....	22
Cuadro 5. Estadígrafos de propiedades de suelos usados en la construcción de la matriz de proyección de parámetro externo de ortogonalización (EPO).....	27
Cuadro 6. Medidas de ajuste (PRIQ, S, RMSE) para predecir pH, arcillas y cationes de intercambio (Ca^{+2} , Mg^{+2} , K^{+} y Na^{+}) con las muestras secas (validación) y después de efectuar el EPO mediante el modelo de mínimos cuadrados parciales (PLSR) y modelo cubista (CM). () indica cuál es el nivel de humedad que obtuvo la mejor estimación en los datos de calibración (Cal) y validación (Val).....	30

ÍNDICE DE FIGURAS

Página

Figura 1. Diagrama esquemático del análisis de imágenes hiper-espectrales, adaptado de Malmir <i>et al.</i> (2019)	7
Figura 2. Diagrama de flujo para el parámetro externo de ortogonalización (EPO) desarrollado en este estudio. Adaptado de Wijewardane <i>et al.</i> (2016)	10
Figura 3. Probabilidad acumulada de la distribución de χ^2 en relación con la distancia de Mahalanobis. Las cruces rojas indican valores fuera de tipo. La línea vertical es el quintil de corte al 97,5%.....	15
Figura 4. Reflectancia de los espectros utilizados en este estudio.....	16
Figura 5. Espectros recortados con un total de 52 longitudes de onda.....	17
Figura 6. Ajuste de Savitsky-Golay a los espectros.....	18
Figura 7. Absorbancia de los espectros 1, 2, 3 y 6 días luego de la hidratación para construir el EPO (arriba)[a] y absorbancia de los espectros 1, 2, 3 y 6 luego de aplicar EPO sobre los mismos (abajo)[b] para el espectro 15.....	25
Figura 8. Proyección de la matriz P resultante del EPO. Los ejes son los índices de las ondas.....	26
Figura 9. Predicción de pH, contenido de arcillas y cationes de intercambio (Ca^{+2} , Mg^{+2} , K^{+} y Na^{+}) en el set de calibración (lado izquierdo) y en el set de validación (lado derecho) luego de corregir con EPO mediante modelo cubista (CM).....	29
Figura 10. Predicción de pH, contenido de arcillas y cationes de intercambio (Ca^{+2} , Mg^{+2} , K^{+} y Na^{+}) en el set de calibración (lado izquierdo) y en el set de validación (lado derecho) luego de corregir con EPO mediante modelo de mínimos cuadrados parciales (PLSR).....	32

RESUMEN

El uso de imágenes hiper-espectrales para determinar propiedades del suelo, permite obtener una mayor cantidad de información en poco tiempo y a un menor costo. La espectrometría de infrarrojo cercano (NIR) ha ido evolucionando en los últimos años en equipos de menores tamaños y más accesibles. En este trabajo se calibró un espectrómetro portátil de infrarrojo cercano marca NeoSpectra-Módulo 2.5, que mide longitudes de onda en el rango de 1.350 nm a 2.500 nm, para la estimación del pH, porcentaje de arcilla y cationes de intercambio. Para calibrar los datos obtenidos por el equipo con los de laboratorio (n= 160), se utilizaron los modelos de mínimos cuadrados parciales (PLSR) y el modelo cubista (CM). La exactitud de los modelos se determinó por el R^2 (coeficiente de determinación) y la raíz cuadrada media del error (RMSE) entre otros. Además, se utilizó una matriz de corrección parámetro externo de ortogonalización (EPO), que elimina el efecto del contenido de agua del suelo. Como resultados se obtuvieron predicciones no muy satisfactorias para el pH y contenido de arcilla ($R^2 < 0,2$), mientras que en la estimación de cationes de intercambio los resultados fueron variables, con R^2 y RMSE en el rango 0,07 a 0,51 y 0,40 y 8,02, respectivamente. Una mejor predictibilidad se obtuvo al utilizar todo el ancho de bandas que mide el equipo. Al utilizar EPO se pudo disminuir el RMSE en un máximo del 78% al utilizar PLSR o CM, permitiendo disminuir el error experimental generado por el efecto del agua en el suelo. El uso de un espectrómetro permite evaluar estas propiedades con una baja predictibilidad al compararse con la metodología de laboratorio. El aumento del número de muestras utilizadas en la calibración, variar el procesamiento de datos o evaluar equipos de espectrometría que posean mayor resolución, podría mejorar las estimaciones.

Palabras claves: Espectrometría, pH, arcilla, cationes de intercambio, agricultura digital.

ABSTRACT

Spectrometric images use for soil properties estimations, allows a more information in short time and low cost. The Near-Infrared (NIR) spectrometry has been developed in the last years with a little and less expensive equipment. pH, clay content and exchangeable cations (Ca^{+2} , Mg^{+2} , K^+ , Na^+) predictions are important as indicators of agricultural management and soil fertility. In this work a NIR spectrometer NeoSpectra Module 2.5 is calibrated in a range 1.300-2.500 nm for estimating these properties. For calibrated these data with laboratory test (n=160), partial least square regression (PLSR) and cubist model (CM) were applied. The model performance is determined with a determination coefficient (R^2) and root mean square error (RMSE) among others. Additionally, an external orthogonalization parameter (EPO) was utilized for moisture sample correct. The performance of pH and clay content prediction was not successful ($R^2 < 0,2$), while cation exchange variable prediction was with R^2 and RMSE from 0.07 to 0.51 and from 0.40 to 8.02, respectively. A best predictability was obtained utilizing all wavelengths that equipment provided. The EPO application decreases the RMSE at maximum 78% when PLSR o CM is utilized, reducing the moisture error sample. The spectrometer predicts these properties with a low predictability in comparison with a laboratory method. A calibration with a major sample number, change in data processing and resolution equipment could to improve estimations.

Keywords: Spectrometry, pH, clay, exchangeable cations, digital agriculture.

INTRODUCCIÓN

Propiedades de suelo y técnicas pedométricas

El suelo como recurso natural tiene una naturaleza heterogénea (McBratney *et al.*, 2019). Esto hace que sus propiedades intrínsecas cambien incluso en una escala espacial de menos de un metro. Por ello, para realizar estudios científicos en este recurso se debe contar con una gran cantidad de datos, que deben ser recolectados y analizados. Sin embargo, acceder a esa información resulta en un número alto de horas-hombre, a través de estudios prolongados en el tiempo y con un costo alto (Tahmasbian *et al.*, 2018; Malmir *et al.*, 2019). Aunque una manera de poder simplificar dichas complejidades podría realizarse a través de la observación del suelo y asociarlo a polígonos que agrupen características similares, no es posible cubrir todas las diferencias existentes (McBratney *et al.*, 2019).

Para poder dar solución a estas problemáticas nace la pedometría a comienzos de la década del 90 (McBratney *et al.*, 2019). Esta rama de la Ciencia del Suelo provee, mediante la aplicación de métodos cuantitativos y la geoestadística, formas más fáciles y de bajo costo para recolectar datos, logrando de esta manera una eficiencia mayor en los muestreos. Lo anterior permite obtener una calidad mayor de mapas de suelos como herramienta para entender de mejor manera las propiedades de suelo y su comportamiento (McBratney *et al.*, 2019).

Desde ese hito, el interés por esta rama de la Ciencia del Suelo se ha ido incrementando en el tiempo (McBratney *et al.*, 2019) lo que, sumado al avance tecnológico, ha hecho que los mapeos de suelos que antes tomaban meses ahora tomen algunos días. El uso de programas estadísticos y computadoras ha permitido obtener imágenes de suelo en áreas más grandes, lo que se explica por una relación sinérgica entre el suelo y la geoestadística, debido a la alta variabilidad del primero (Minasny y McBratney, 2016).

Para la obtención de estas imágenes, desde las cuales se predicen las propiedades del suelo, se utilizan sensores proximales que se colocan a no más de 2 m del suelo, tomando imágenes hiperespectrales que abarcan el rango de luz visible al infrarrojo cercano (NIR, por sus siglas en inglés, Near Infra Red) (Viscarra-Rossel *et al.*, 2011). Estos equipos se basan en principios como: i) las partículas tienen ciertas frecuencias a las cuales vibran o rotan de acuerdo a niveles discretos de energía (Horta *et al.*, 2015; Tahmasbian *et al.*, 2018); ii) utilizan longitudes de onda desde los 350 nm a 2500 nm para explorar la interacción entre la radiación incidente y la reflectancia de un suelo; y iii) la absorción es facilitada principalmente por los grupos funcionales de la materia orgánica (C-H, N-H u O-H) de la matriz del suelo (Chang *et al.*, 2005).

Si bien estas técnicas, al evaluar propiedades de manera indirecta, pueden generar incertidumbre o requerimientos de calibración, presentan la gran ventaja que se puede tomar una gran cantidad de muestras en poco tiempo y no son destructivos (Viscarra-Rossel *et al.*,

2011). Adicionalmente, los sensores actuales son más pequeños, rápidos y baratos que los de primera generación, siendo igual o más eficaces que sus versiones previas (Tahmasbian *et al.*, 2018; Malmir *et al.*, 2019; Barthes *et al.*, 2019).

En la actualidad las tendencias apuntan a que un equipo por si solo es capaz de determinar propiedades de suelo, dejando de lado el uso de equipos complementarios para estimarlos (Wang *et al.*, 2015, Tahmasbian *et al.*, 2018, Sharififar *et al.*, 2019). Si bien el parámetro que se ha predicho más exitosamente es el contenido de carbono orgánico del suelo (COS), otras propiedades tales como el pH, el contenido de arcillas y los cationes de intercambio como los iones calcio (Ca^{+2}), magnesio (Mg^{+2}), potasio (K^{+}) y sodio (Na^{+}) se han podido estimar exitosamente (Minasny *et al.*, 2011; Lin *et al.*, 2018; Malmir *et al.*, 2019; Ng *et al.*, 2020; Lobos *et al.*, 2021; Sepulveda *et al.*, 2021).

El rango del espectro electromagnético utilizado para la determinación de pH, contenido de arcilla y cationes intercambio han sido objeto de estudio, sobre todo en equipos portátiles modernos que trabajan con rangos de longitud de ondas más acotados (Barthes *et al.*, 2019). Malmir *et al.* (2019) encontraron que, en la determinación de algunos nutrientes como boro (B) y Ca, el uso de longitudes de onda desde 400-1.000 nm es más que suficiente para realizar una predicción de su contenido en el suelo, mientras que para K se necesita un rango más amplio (Xu *et al.*, 2018). Estos rangos a su vez se encuentran relacionados a la determinación de nitrógeno (N), carbono (C), así como a otros nutrientes como cinc (Zn), Mg, Ca y cobre (Cu) (Stenberg *et al.*, 2010; Gmur *et al.*, 2012; Tahmasbian *et al.*, 2018; Malmir *et al.*, 2019).

Por otra parte, la exactitud de la medición realizada podría estar condicionada por el tamaño de las partículas, donde un menor tamaño de partícula aumenta la superficie específica (Stenberg *et al.*, 2010); con esto se podría aumentar la reflectancia y subestimar el valor comparado con una partícula más gruesa (Cooper y Mustard, 1999). Sin embargo, en estudios recientes no se ha visto diferencias en las mediciones logradas entre suelos tamizados y pulverizados (Tamburini *et al.*, 2017; Malmir *et al.*, 2019). Adicionalmente, el contenido de agua de la muestra se ha señalado como una variable que podría distorsionar los resultados, causando una estimación pobre de otras propiedades como COS (Minasny *et al.*, 2009; Morgan *et al.*, 2009; Minasny *et al.*, 2011) o contenidos de arcilla (Ge *et al.*, 2014) siendo este factor uno de los más importantes al momento de hacer determinaciones en terreno (Wijewardane *et al.*, 2016). Finalmente, otro de los factores que podría estar afectando la medición es la temperatura, el cual se puede remover mediante el uso de un parámetro externo de ortogonalización (EPO) (Roger *et al.*, 2003). Sin embargo, si la medición se realiza en condiciones controladas de laboratorio su efecto se podría verse reducido (Ge *et al.*, 2014).

El EPO ha emergido como un método eficaz para remover el efecto del contenido de agua del suelo del espectro del NIR. Este parámetro externo descompone el espectro en dos partes, una útil que tiene directa relación con la variable respuesta, y otra que se está influenciada por agentes externos o parasítica (Wijewardane *et al.*, 2016). Su uso se ha trasladado desde la eliminación de la temperatura para medir contenido de azúcar en frutos (Roger *et al.*, 2003), hasta para ajustar la estimación de COS y contenido de arcilla (Minasny *et al.*, 2011; Ge *et al.*, 2014; Liu *et al.*, 2019).

Se han usado diferentes modelos estadísticos para calibrar los valores de referencia y los otorgados por el equipo, siendo estos de origen lineal o no lineal (Cardelli *et al.*, 2017; Sharififar *et al.*, 2019). La regresión de mínimos cuadrados parciales, el árbol de regresión cúbica y bosque aleatorio son algunos de los modelos que se utilizan, los cuales muestran mejores ajustes dependiendo de las diferentes situaciones experimentales, de manera individual como en conjunto (Wang *et al.*, 2015; Malmir *et al.*, 2019; Sharififar *et al.*, 2019). Conjuntamente, la transformación de datos previa al ingreso del modelo también mejora el grado de ajuste, disminuyendo el ruido experimental con relación a datos no procesados o sin transformación (Malmir *et al.*, 2019; Sharififar *et al.*, 2019).

Importancia de la estimación de pH, contenido de arcilla y cationes de intercambio en los suelos

El pH del suelo se puede definir estrictamente como el logaritmo negativo de la actividad del ion hidrógeno (H^+) en la solución suelo (Sadzawka, 2006). Sin embargo, desde el punto de vista agronómico provee el índice de disponibilidad y reserva de nutrientes en un suelo determinado en cuanto a su contenido de cationes (Hirzel, 2014). Lo anterior responde a que, por ejemplo, en un suelo de pH más alto predominarán los cationes tales como Ca^{+2} , Mg^{+2} , K^+ , Na^+ y estarán en menor proporción el aluminio (Al^{+3}) y H^+ (Sadzawka, 2006).

La distribución de cationes de intercambio es importante y propia de cada suelo, y su correcto balance permite el establecimiento de frutales y/o cultivos supliendo sus necesidades nutricionales; por el contrario, si estas están desbalanceadas deben ser corregidas (Hirzel, 2014). La importancia de estos cationes radica en que participan de la rigidez de la pared celular (Ca), la formación de la molécula de clorofila (Mg) y el co-transporte de azúcares hacia los frutos (K) entre otras funciones (Havlin, 2017).

El contenido de arcilla se relaciona directamente con la fertilidad del suelo, debido tanto a su capacidad de generar carga, como por su alta reactividad que permite que los cationes puedan ser adsorbidos a su superficie, liberándose poco a poco de manera de equilibrar la solución suelo (Sadzawka, 2006). La combinación de estas tres variables podría dar algunas ideas sobre cómo puede ser la fertilidad de un suelo, por ejemplo: en un mismo valor de pH, un suelo con mayor contenido de arcilla tendrá una mayor suma de cationes de intercambio versus uno con una dominancia de las partículas de tamaño arena (Hirzel, 2014).

La determinación de estos parámetros en un análisis de suelo requiere una gran cantidad de reactivos. Considerando lo anterior, más el tiempo empleado en coleccionar las muestras y la entrega de resultados hace que el uso de equipos de espectrometría NIR, sea una buena alternativa para obtener estos parámetros de manera rápida y efectiva para la toma de decisiones.

HIPÓTESIS

El pH, el contenido de arcilla y los cationes de intercambio pueden ser estimados satisfactoriamente mediante el uso de un espectrómetro de infrarrojo cercano (NIR), en suelos bajo manejo agrícola en el valle longitudinal de Chile central.

OBJETIVOS

Objetivo general

Validar un espectrómetro de infrarrojo cercano portátil para la determinación de pH, contenido de arcilla y cationes de intercambio.

Objetivos específicos

- Evaluar el pH, contenido de arcilla y cationes de intercambio en un amplio rango de suelos y manejos en Chile central, mediante un espectrómetro de infrarrojo cercano y métodos de laboratorio.
- Seleccionar el mejor ajuste para reducir el error experimental generado por el contenido de agua en el suelo en las estimaciones de pH, contenido de arcilla y cationes de intercambio por un espectrómetro de infrarrojo cercano portátil.

MATERIALES Y MÉTODOS

Área de estudio

El estudio abarcó un transecto de 190 km del valle longitudinal de Chile continental entre las Regiones Metropolitana y del Maule (paralelos 33° y 35°). La zona incluyó una gran variedad de tipos de suelos y climas, que abarcan diversos sistemas agropecuarios y otros ecosistemas descritos por Pfeiffer *et al.* (2020) como: bosque de hoja caduca, bosque de hoja ancha, bosque esclerófilo y bosque espinoso.

Toma de muestras

Las muestras fueron obtenidas desde distintos sitios ubicados en las regiones Metropolitana, O'Higgins y del Maule, totalizando 170 muestras colectadas.

Las muestras de suelo fueron colectadas en un intervalo de 0-30 cm de profundidad, mediante el uso de una pala retirando la capa superficial del suelo totalizando un total de 1 kg por cada sitio muestreado. Posteriormente el lugar fue georreferenciado mediante un GPS (latitud, longitud y elevación). Además, junto a cada muestra de suelo, se colectó una muestra con cilindro metálico entre los 10-20 cm de profundidad para determinar la densidad aparente del suelo (Da) mediante el método propuesto por Sandoval *et al.* (2012).

Las muestras fueron clasificadas de acuerdo a su régimen de suministro de agua (secano o riego), la especie vegetal precedente (cultivo o frutal) y otras prácticas de manejo (labranza, quema o incorporación de rastrojo) según corresponda a cada caso. Adicionalmente, fueron clasificadas por su orden, régimen de humedad y temperatura de los suelos, de acuerdo con Casanova *et al.* (2013).

Calibración y análisis estadístico

Las muestras de suelo fueron trasladadas al Laboratorio de Química de Suelos y Aguas, pertenecientes al Departamento de Ingeniería y Suelos de la Facultad de Ciencias Agronómicas de la Universidad de Chile.

Cada muestra fue secada al aire, tamizada a 2 mm y dividida en dos submuestras. A una de ellas se le realizó la determinación de:

pH en agua. El pH se midió en agua usando un pHmetro (Hanna Instruments) con electrodo combinado. La medición consideró una relación suelo:agua 1:2,5 (Sadzawka *et al.*, 2006).

Contenido de arcilla. Se determinó colocando una muestra de 50 g de suelo tratado previamente con polifosfato de sodio como dispersante en una probeta de 1 L, conocido como método del hidrómetro de Bouyoucos (Sandoval *et al.*, 2012).

Cationes de intercambio. Se determinó con solución de acetato de amonio 1 M a pH 7,0 según Sadzawka *et al.* (2006).

Desde la otra submuestra se recolectó en promedio 30 g de suelo, la cual se distribuyó uniformemente en una placa de Petri. En ella se midió la reflectancia de la muestra de suelo con un equipo portátil de bajo costo de reflectometría de rojo cercano marca NeoSpectra-Modulo 2.5, que mide longitudes de onda desde 1.350 nm a 2.500 nm, el cual entrega un valor de reflectancia corregida (R). Este valor adimensional, se obtuvo mediante una calibración previa con una superficie blanca provista por el propio equipo (Figura 1).

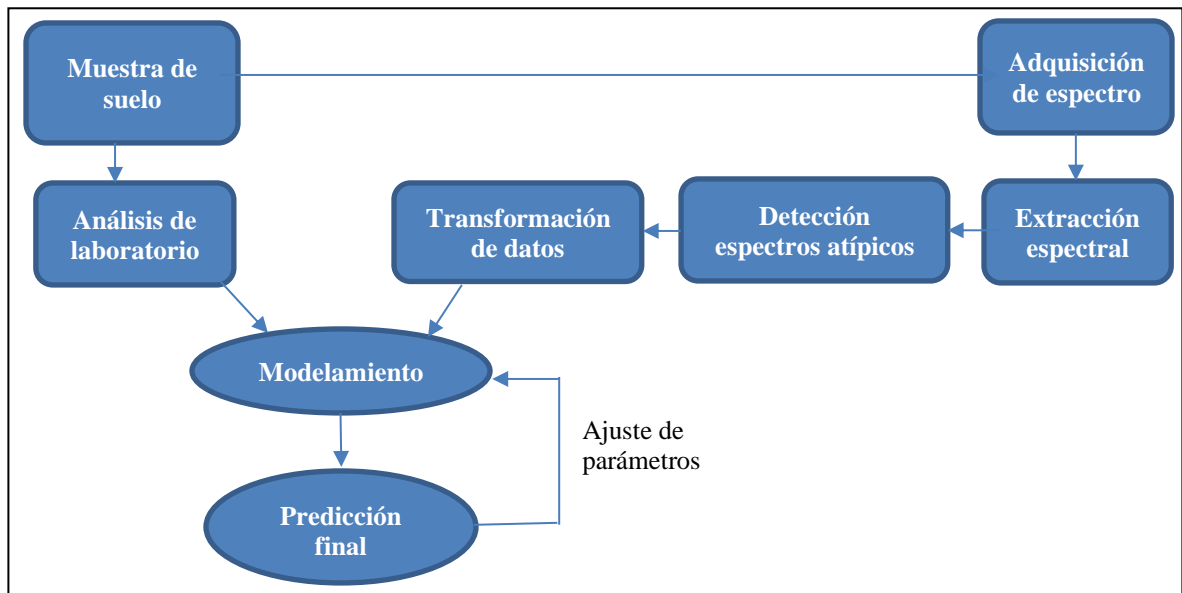


Figura 1. Diagrama esquemático del análisis de imágenes hiper-espectrales adaptado de Malmir *et al.* (2019).

Una vez obtenida la R se procedió a detectar los valores atípicos por medio de la metodología propuesta por Filzmoser *et al.* (2005). Para ello, se comenzó con un análisis de componentes principales, con el que se calculó el promedio y la covarianza de los primeros cinco componentes. Posteriormente, se calculó la distancia de Mahalanobis de acuerdo con la ecuación 1:

$$MD_i = ((x_i - t)^T C^{-1} (x_i - t))^{1/2} \text{ para } i = 1, \dots, n. \quad (\text{Ec. 1})$$

donde:

x_i = valor del componente

t = promedio aritmetico multivariado

C = matriz de covarianza

T = matriz transpuesta

Los datos de reflectancia presentaron una distribución normal, por lo que, los valores de la distancia de Mahalanobis al cuadrado (MD_i^2) fueron asociados a una distribución de chi-cuadrado (χ_p^2) (Filzmoser *et al.*, 2005). Por lo tanto, se trabajó con una distribución de χ_p^2 con 5 grados de libertad (p) y con una probabilidad del 97,5% ($1-\alpha$). Considerando que los espectros presentan la misma distribución, se determinó el valor crítico (p_{crit}) para distinguir entre los valores atípicos y aquellos en el límite que pueden ser usados en la simulación de acuerdo con la ecuación 2:

$$p_{crit}(\partial, n, p) = \frac{0,24 - 0,003p}{\sqrt{n}}, \text{ para } p \leq 10 \quad (\text{Ec. 2})$$

donde

n = número de variables (longitudes de onda)

p = grados de libertad

∂ = distribución $\chi_{p;1-\alpha}^2$

A partir de este análisis se procedió a descartar los espectros que estaban fuera del valor crítico estimado. Posterior a eso, los valores de reflectancia fueron transformados a absorbancia mediante la ecuación 3:

$$\text{absorbancia} = \log\left(\frac{1}{\text{reflectancia}}\right) \quad (\text{Ec. 3})$$

Con estos valores transformados, se procedió a seleccionar solo las bandas útiles para el estudio. Luego, estas fueron suavizadas usando el algoritmo de Savitzky-Golay (SG) con una ventana de 11 datos y una regresión polinomial de grado 3. Para remover el efecto de base, se utilizó la transformación variada normal estándar (SNV), donde cada espectro fue normalizado tomando su promedio y dividiéndolo por su desviación estándar (Minasny *et al.*, 2011).

Determinación del error por contenido de agua de la muestra (EPO)

Se determinó el efecto del contenido de agua del suelo sobre la exactitud de la predicción. Para ello, del total de muestras recolectadas, se formaron dos grupos distintos los cuales se detallan a continuación:

- Grupo 1: Datos para desarrollo del parámetro de ortogonalización externo (EPO). Se seleccionaron 39 muestras del total de datos que representaron todo el espectro de contenidos de pH, contenido de arcilla y cationes de intercambio.
- Grupos 2: Datos de calibración. El resto de los datos (121 muestras) fueron usados para el modelo de calibración, relacionando el espectro con el contenido de pH, contenido de arcilla o cationes de intercambio, usando los modelos de mínimos cuadrados parciales (PLSR) y/o el modelo cubista (MC)

De las muestras del grupo 1, fueron colocados 30 g de suelo en una placa Petri con un espesor de 5 mm aproximadamente y se hidrataron con agua destilada de manera que quedaron al límite de adherencia o como pasta saturada, de acuerdo Minasny *et al.* (2011). Luego las muestras fueron cubiertas con su respectiva tapa y dejadas para que se saturaran completamente por dos días. Después de este período, se les realizó una fotografía con el equipo, fueron masadas volviéndolas a dejar cerradas en condiciones frescas y secas, labor que se repitió por 6 días. Luego, los espectros fueron sometidas al procedimiento EPO.

El EPO asume que la matriz espectral X puede ser descompuesta utilizando la ecuación 4:

$$X = XP + XQ + R \quad (\text{Ec. 4})$$

donde XP es la parte útil del espectro, XQ es la parte no útil (influenciada por el contenido de agua) y R el ruido provocado por la falta de ajuste. El parámetro XP se expresó como $X(I-Q)$, donde I es la matriz de identidad, en tanto que Q puede ser escrita como $Q = GG^T$, donde (T) es la matriz transpuesta.

Para determinar G , la parte no informativa del espectro, que es ortogonal a la parte útil del espectro, Roger *et al.* (2003) sugirieron usar el componente principal de la diferencia de espectro D , que es una matriz asimétrica (tamaño $n \times m$). Este valor de D es la diferencia entre el espectro con el tratamiento con agua y el seco al aire, como se muestra en la ecuación 5:

$$D = [X_{ad}] - [X_{d1}] \quad (\text{Ec. 5})$$

donde X_{ad} es el espectro a condiciones de secado al aire y X_{d1} es el espectro al primer después de rehidratar.

El algoritmo de EPO se realizó de acuerdo con Minasny *et al.* (2011), como se muestra:

1. Calcular diferencia espectro D (tamaño $n \times m$): $D = X_{\text{húmedo}} - X_{\text{seco}}$.
2. Mejorar un análisis de componente principal sobre $D^T D$. Esto se obtiene mediante la descomposición de $D^T D$ para obtener USV^T , donde U es una matriz $n \times n$, S es una matriz diagonal $n \times m$ y V es una matriz $m \times m$.
3. Definir el número c de factores de EPO, y obtener V_s , un subconjunto de m por c de V (G).
4. Estimar Q de GG^T , ya que $Q = GG^T$.
5. Finalmente se calcula la matriz $P = I - Q$ (donde P es una matriz simétrica), calculándose los espectros transformados como $X^* = XP$ (Figura 2).

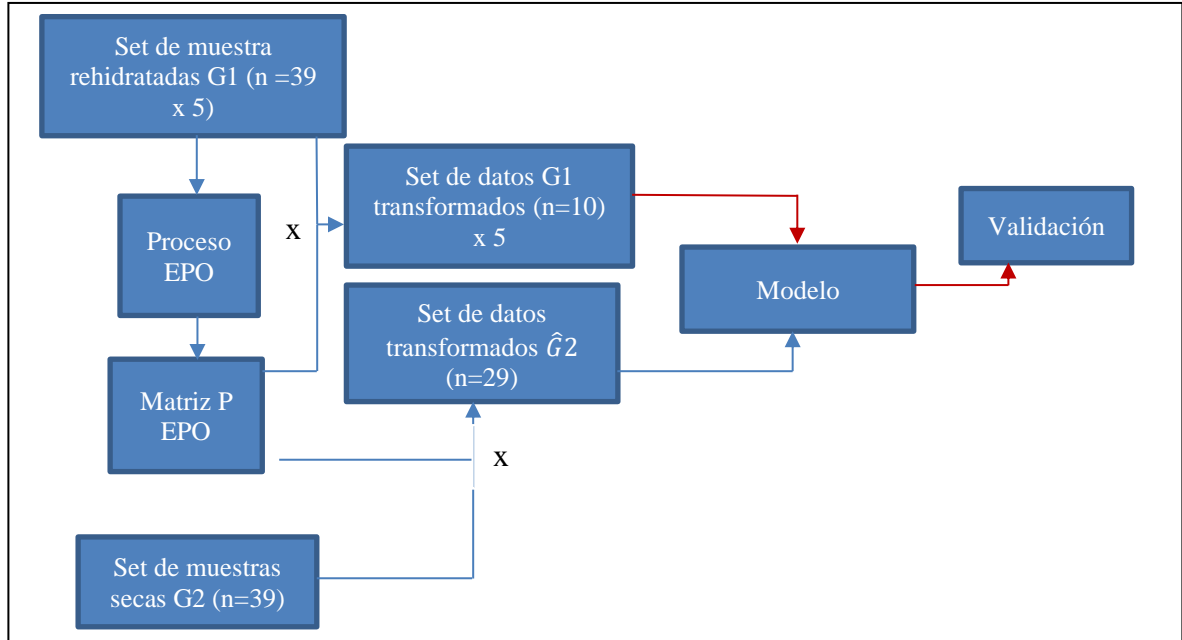


Figura 2. Diagrama de flujo para el parámetro externo de ortogonalización (EPO) desarrollado en este estudio. Adaptado de Wijewardane *et al.* (2016).

El único parámetro que debió ser elegido en EPO fue c , que es el número de dimensiones del componente principal de DD^T , el cual representa el espacio inútil que necesita ser removido. Esto se determinó mediante el uso del parámetro de Wilks' Λ , que define qué tan similar fue un espectro transformado de dos o más niveles de contenido de agua (variación intergrupala), comparado con la diferencia entre dos espectros de dos muestras diferentes (variación intra-grupal) (Roger *et al.*, 2003). Para ello se propuso la ecuación 6:

$$\text{Wilk's } \Lambda = \text{tr}(B)/\text{tr}(T) \quad (\text{Ec. 6})$$

donde B es la variación intergrupala, mientras que T es la variación intra-grupal.

Posteriormente, se graficó este valor con relación al número de componentes principales desde 1 a 20. Se consideró que el número de componente principal es adecuado desde cuando el valor de Wilk's $\Lambda=1$ (Wijewardane *et al.*, 2016).

Calibración de muestras

Para calibrar los datos obtenidos por medio del equipo con los de laboratorio, se utilizaron los modelos de mínimos cuadrados parciales (PLSR) y el modelo cubista (MC). La exactitud de los modelos se determinó por el R^2 (coeficiente de determinación) (ecuación 7), la raíz cuadrada media del error (RMSE) (ecuación 8), el sesgo (S) (ecuación 9) y el radio de mejoramiento de distancia entre cuartiles (RPIQ) (ecuación 10) (Minasny *et al.*, 2011; Wadoux *et al.*, 2021).

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{Ec. 7})$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{Ec. 8})$$

$$S = \frac{\sum_{i=1}^n \hat{y}}{n} - \frac{\sum_{i=1}^n y}{n} \quad (\text{Ec. 9})$$

$$\text{RPIQ} = \frac{(Q_3(y) - Q_1(y))}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2}} \quad (\text{Ec. 10})$$

donde \hat{y} e y son los valores predichos y los de referencia, respectivamente. Q_3 y Q_1 son el primer y el tercer quintil, respectivamente.

Para la calibración de los datos en seco éstos se separaron en dos, uno de ellos de llamado de calibración (75% del total de los datos) el cual se usó en el modelamiento con PLSR y CM, y otro de validación (25% del total de datos) que se usó para evaluar el desempeño del modelo realizado previamente. El modelo seleccionado fue aquel que mediante el ajuste de los componentes principales obtuvo el menor RMSE, ME y el R^2 .

En el caso de la calibración mediante EPO se procedió de la siguiente manera (Figura 2):

1. Se seleccionaron los mismos espectros provenientes de las muestras secas que aquellas que fueron usados para crear la matriz P (39 espectros). Desde ese conjunto se utilizó un 75% de las muestras secas ponderadas por la matriz EPO y se calibró un modelo usando mínimos cuadrados parciales (PLSR) y modelo cubista (MC).
2. Para validar el modelo, se utilizó el 25% restante del conjunto de espectro de las muestras con diferentes niveles de humedad (EPO1, EPO2, EPO3, EPO5 y EPO6) ponderados por la matriz P. Resultando un total de 5 validaciones distintas por cada variable respuesta.
3. La mejor validación resultó ser aquella que tuvo el menor RMSE y el mayor R^2 según fuese el caso, dependiendo del nivel de humedad usado (EPO1, EPO2, EPO3, EPO5 y EPO6).
4. Para mejorar el grado de ajuste, se modificaron los parámetros de cada modelo. En el caso de PLSR se modificó el número de componentes principales usados en la modelación (CP), mientras que en el caso del modelo cubista se modificó el número de iteraciones (*committees*) a realizar.

Se utilizó el lenguaje de programación R versión 3.6.1 para el análisis de los datos.

RESULTADOS Y DISCUSIÓN

Características de las muestras de suelo

Un total de 170 muestras fueron caracterizadas para este estudio. Las muestras recolectadas en la Región de O'Higgins fueron de 126, mientras que en Maule y Metropolitana fueron de 35 y 9, respectivamente. Las muestras de suelo fueron en su totalidad de sectores bajo riego, mientras un 91% se dedicaba a la producción frutal, un 7% se usaba para el cultivo de hortalizas y el resto fueron suelos en barbecho y/o praderas naturales. Las propiedades evaluadas de estos suelos se presentan en el Cuadro 1.

Cuadro 1. Propiedades y estadígrafos de los suelos utilizados en el estudio.

Propiedad	Estadígrafos ¹			
	X	Mín	Máx	CV
pH (-)	6,6±0,56	4,75	8,03	8,53
Contenido de arcilla (%)	26,1±8,80	6,20	50,6	34,00
Ca ⁺² intercambiable (cmol ₍₊₎ kg ⁻¹)	17,6±10,10	1,92	55,9	57,20
Mg ⁺² intercambiable (cmol ₍₊₎ kg ⁻¹)	2,5±1,30	0,03	6,61	51,70
Na ⁺ intercambiable (cmol ₍₊₎ kg ⁻¹)	2,8±3,10	0,12	29,9	100,90
K ⁺ intercambiable (cmol ₍₊₎ kg ⁻¹)	0,69±0,49	0,06	2,8	72,00

¹ X = promedio; ± = desviación estándar ; Mín = mínimo; Máx = máximo; CV = coeficiente de variación (%).

El pH de los suelos varió en un rango desde lo que se considera como muy fuertemente ácido (4,75) hasta moderadamente alcalino (8,30) (Sadzawka *et al.*, 2006). Los contenidos de arcillas indicaron que las clases texturales de los suelos variaron entre franco arenosa y arcillo limosa representando un 7% y 2,3% de las muestras totales, respectivamente. Las clases texturales predominantes fueron franca y francas finas (FAa, FAL, FA y FL) representando un 41% y 45,8%, respectivamente.

El contenido de cationes de intercambio se asemeja a aquellos definidos por Luzio *et al.* (2010) para algunos suelos de las regiones de O'Higgins y Maule. En particular, estos fueron asociados a sectores donde predominan pendientes de 3 al 5% con contenidos de Ca⁺², Mg⁺², K⁺ y Na⁺ intercambiables de 16,8, 3,0, 0,4 y 0,4 cmol₍₊₎ kg⁻¹, respectivamente, como por ejemplo la Serie de Suelo Huecán. La suma de cationes promedio de las muestras de este estudio fue cercana a 24 cmol₍₊₎ kg⁻¹, la cual también coincide con el pedón característico de la zona (Luzio *et al.*, 2010). La relación de los cationes en el complejo de intercambio también fue analizada, de forma que los porcentajes de Ca⁺, Mg⁺², K⁺ y Na⁺ intercambiables en la suma de cationes fueron de 72±10,0, 11±5,6, 3,5±2,9 y 12,6±7,6, respectivamente. En este sentido, una proporción de 80%, 12%, 3% y 2% se sugiere como un balance correcto para la nutrición de cultivos y frutales (Hirzel, 2014). Existen al menos 3 muestras que presentaron valores de saturación de Na⁺ intercambiable sobre 35%, lo cual permitiría inferir

que se está frente a suelos sódicos. Sin embargo, se requieren de análisis adicionales como la capacidad de intercambio catiónico, la relación de adsorción de sodio (RAS) y el porcentaje de saturación de sodio (PSI) para su determinación (Sadzawka, 2006b).

De acuerdo con la georreferenciación realizada y las descripciones publicadas por CIREN (1996, 1997 y 2002), las muestras de los suelos recolectadas en la Región de O'Higgins pertenecen en un 57% al Orden Mollisol, seguido por Inceptisoles con un 34,6% de las muestras de la Región. En la Región del Maule un 54% fue clasificado como Inceptisol y un 17% como Andisol. Un total de 8 de las 9 muestras de la Región Metropolitana fueron clasificadas como Mollisoles. Otros Órdenes de importancia fueron los Alfisoles con un 4,7% y 8,6% de las muestras de las Regiones de O'Higgins y Maule, respectivamente.

En el caso de los suelos pertenecientes al Orden Andisol en la Región del Maule, estos correspondieron a las Series de suelo Romeral y Cumpeo. Aunque estos suelos son derivados de cenizas volcánicas holocénicas, han tenido una evolución de tal magnitud que ya no podrían ser considerados Andisoles (Luzio *et al.*, 2010). Por ejemplo, en este estudio la densidad aparente (D_a) de estos suelos fue mayor a $0,9 \text{ Mg m}^{-3}$ llegando hasta valores de $1,16 \text{ Mg m}^{-3}$ por lo que no podrían ser considerados en este Orden. Lo anterior, coincide también con otro estudio reportado en la zona por Stolpe *et al.* (2008). Otro indicador es su mineralogía, donde se ha encontrado que, en este caso en particular, predomina la haloisita. Sin embargo, estas determinaciones no fueron objeto de este estudio.

De acuerdo a Luzio *et al.* (2010) los suelos de las regiones de O'Higgins y Maule desarrollados a partir de sedimentos aluviales medios y finos son los más recientes. Si bien su posición es plana, la dispersión de sus propiedades es alta, ya sea en la clase textural como en su profundidad. Lo anterior implicaría que su variabilidad respecto a su clasificación taxonómica también sería alta, desarrollándose suelos clasificados en los Órdenes Alfisol, Mollisol e Inceptisol, coincidiendo la distribución de suelos encontradas en este estudio. A su vez explica la dominancia de clases texturales francas y sus derivados (cercanas al 90%) presentes.

Las muestras colectadas se encuentran dentro de la zona edáfica denominada como Zona Mediterránea Árida, y se encuentran dentro de la zona donde los suelos se caracterizan por tener regímenes de humedad xérico y de temperatura térmico. El régimen xérico corresponde a climas mediterráneo donde la sección de control de humedad se encuentra seca por 45 días o más luego del solsticio de verano, y esta humedad luego de 45 días o más luego del solsticio de invierno. Un régimen de temperatura térmico, en tanto, se caracteriza por una variación anual mayor a $6 \text{ }^\circ\text{C}$ con estacionalidad y con una temperatura media anual de suelo que va desde los $15 \text{ }^\circ\text{C}$ y siendo siempre menor a $22 \text{ }^\circ\text{C}$ (Luzio *et al.*, 2010).

Tratamiento de espectros

El total de espectros de suelos utilizados en este estudio para la calibración, fueron obtenidos de 160 muestras, mientras que 10 espectros de suelos fueron descartados por estar fuera del valor crítico determinado a través de la metodología propuesta por Filmozer *et al.* (2015), como se muestra en la Figura 3.

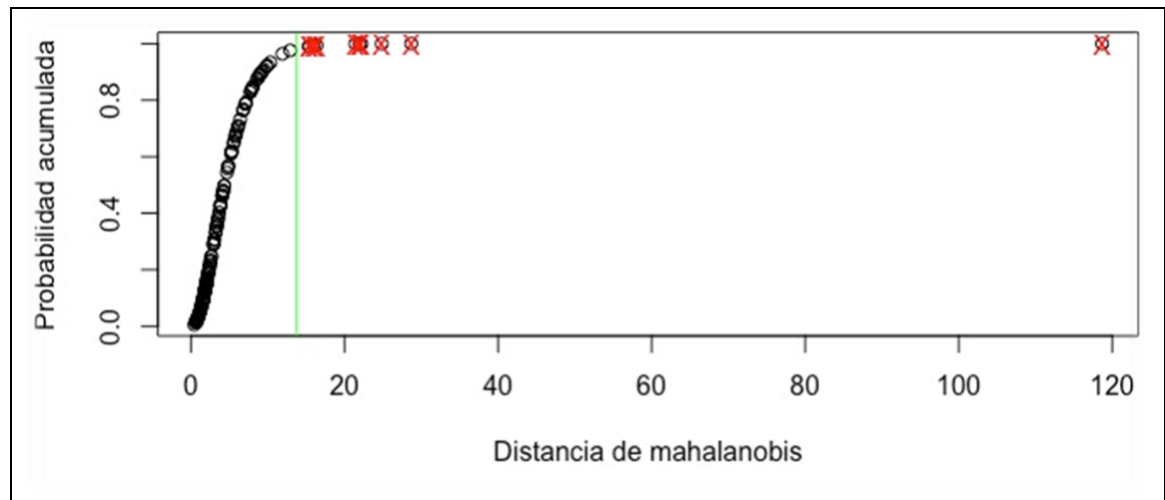


Figura 3. Probabilidad acumulada de la distribución de χ^2 en relación con la distancia de Mahalanobis. Las cruces rojas indican valores fuera de tipo. La línea vertical verde es el quintil de corte al 97,5%.

Los espectros de las muestras de suelos seleccionados presentaron una forma característica, que se manifiesta con una disminución en la reflectancia en el rango de 1.400 nm a los 1.900 nm, así como también cuando alcanza los 2.200 nm (Figura 4). Esto responde en primer lugar al efecto del agua (discutido más adelante), y donde el comportamiento alrededor de los 2.200 nm responde a la presencia de caolinitas u otros filosilicatos posiblemente a los enlaces Al-OH (Nawar *et al.*, 2016; Seema *et al.*, 2020) de sus octaedros. Otro pequeño peak es observado alrededor de los 2.280 nm, el cual se atribuye al Fe-OH, donde el Fe está como catión en los octaedros como por ejemplo en las montmorrillonitas (Nawar *et al.*, 2016). Un pequeño “hombro” a los 2.340 nm puede representar illitas o mezclas de minerales de muscovitas (Nawar *et al.*, 2016).

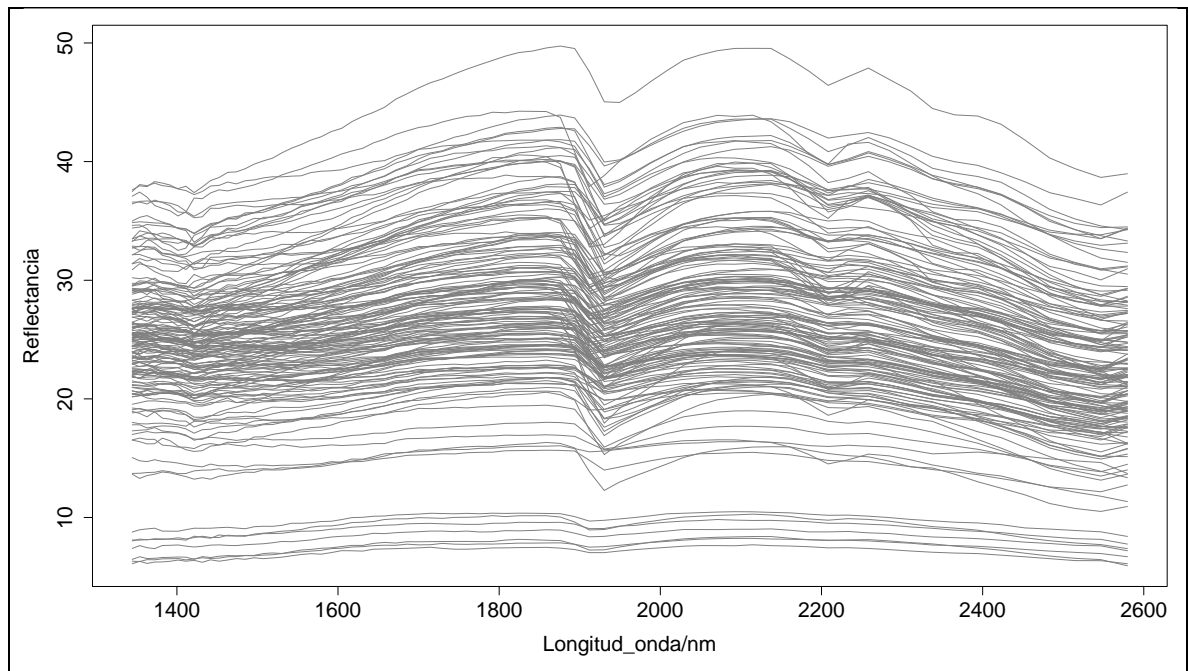


Figura 4. Reflectancia de los espectros utilizados en este estudio.

Cuando los espectros son usados para modelar propiedades de suelo, la mejor forma para trabajarlos es en términos de absorbancia (Gobrecht *et al.*, 2015). De esta forma, los componentes químicos estarán linealmente relacionados a la longitud de onda (Wadoux *et al.*, 2021).

Los espectros obtenidos tuvieron un total de 72 bandas o longitudes de onda desde los 1.250 a los 2.500 nm. El ruido o error en la estimación de alguna propiedad de suelo se encuentra principalmente en los extremos, por lo que se deben eliminar a fin de mejorar el ajuste del modelo. Con este fin se procedió a eliminar diez bandas del extremo superior del espectro y la misma cantidad del otro extremo, resultando en un espectro final de 52 bandas desde los 1.442 a los 2.284 nm. En el extremo superior de la banda, alrededor de los 2.450 nm a los 2.500 nm, la información que se puede extraer es escasa. Además, el rango anterior coincide con el límite de medición del sensor (Wadoux *et al.*, 2021) (Figura 5).

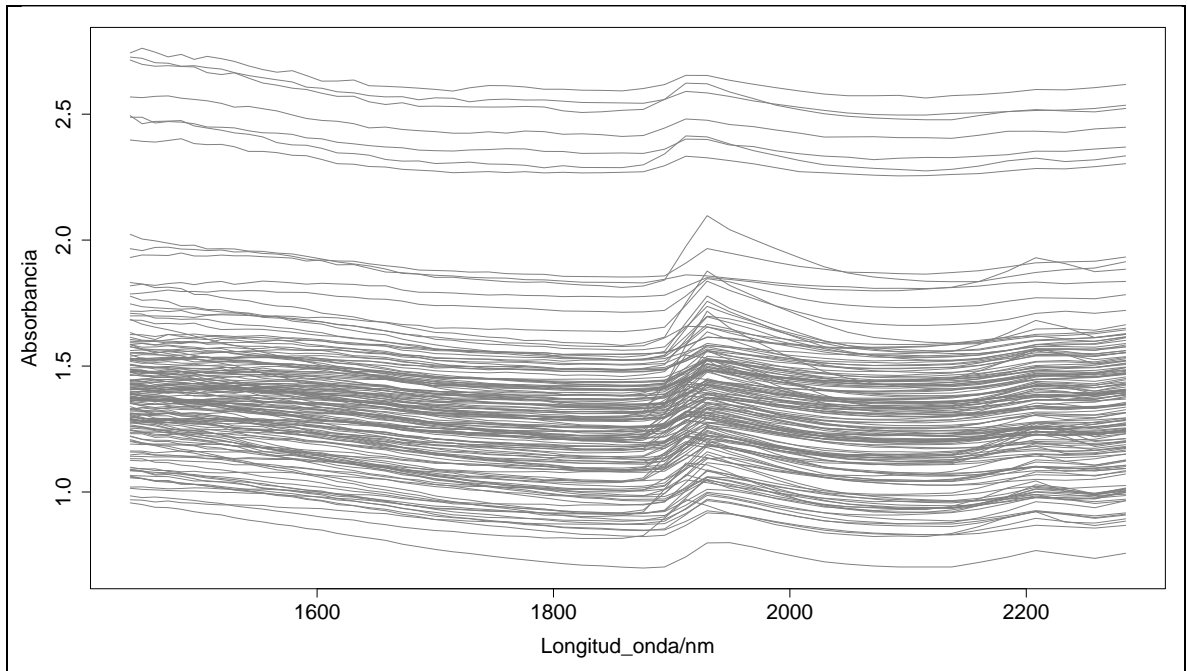


Figura 5. Espectros recortados con un total de 52 longitudes de onda.

La transformación de Savitsky-Golay (SG) permitió ajustar una curva polinomial de tercer grado ($k=3$), que se ajustó adecuadamente a los espectros (Figura 6). El uso de esta herramienta para mejorar la predicción de propiedades de suelo no ha sido del todo útil. Seema *et al.* (2020) encontraron que al estimar COS obtuvieron mejores estimaciones al usar solo el espectro como reflectancia o crudo. Por el contrario, Nawar *et al.* (2016) obtuvieron mejores estimaciones de la MOS al usar SG comparado con un espectro sin tratamiento. Sin embargo, en el caso anterior tuvieron mejores estimaciones en el contenido de arcilla al usar otros tratamientos en el mismo estudio.

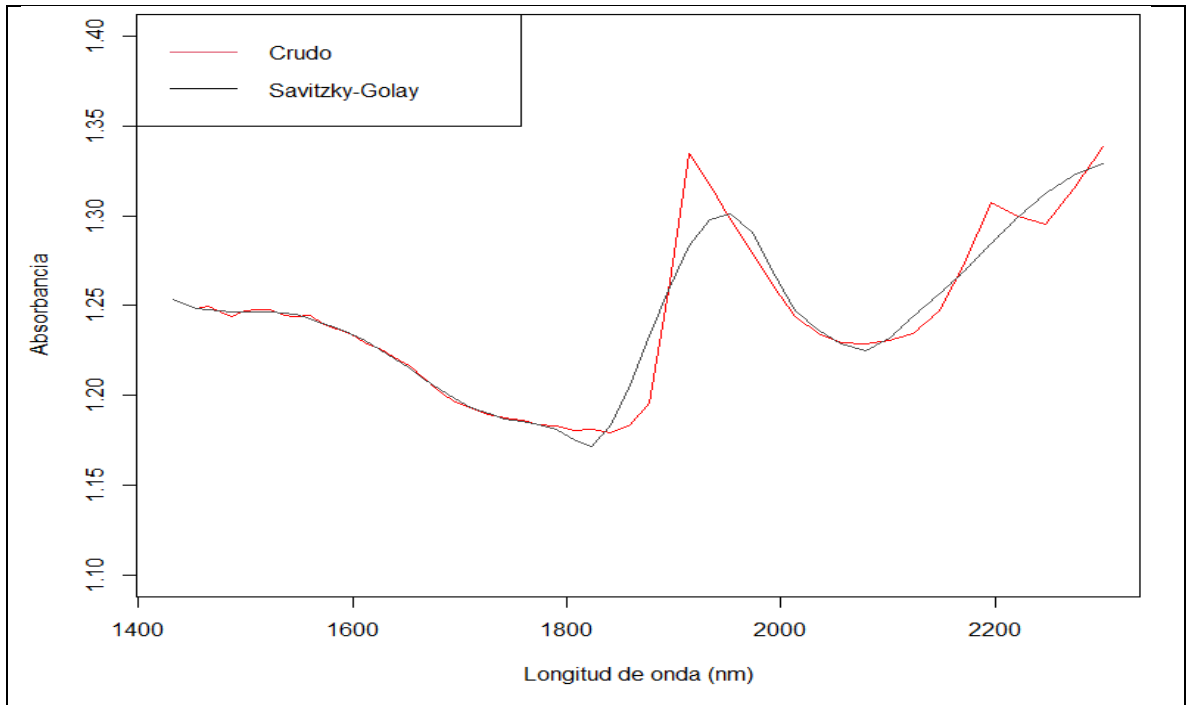


Figura 6. Ajuste de Savitzky-Golay a los espectros de las muestras de suelo.

Finalmente se realizó una corrección por dispersión mediante la variación normal estándar (SNV). Este pre-procesamiento corrige la dispersión de la luz de la muestra de suelo, la cual puede ser atribuida a desviaciones de la longitud de onda, largo del recorrido de la luz como la sensibilidad del detector, siendo SNV el método más sencillo para lograrlo (Siesler *et al.*, 2018; Wadoux *et al.*, 2021).

Calibración de muestras en seco

En los cuadros 2 y 3 se presentan los resultados, luego de aplicar los modelos de mínimos cuadrados parciales (PLSR) y modelo cubista (CM), para la estimación de pH, contenido de arcilla y cationes de intercambio. Para modelar los espectros completos con PLSR, se usaron los primeros 7 componentes principales para la estimación de K y Ca intercambiables, mientras que para el resto de las propiedades se usaron los primeros 5 componentes. Por el contrario, cuando se modeló con los espectros recortados y suavizados, se utilizaron los primeros 5 componentes para estimar todas las propiedades. Para el desarrollo de CM se usaron un total de 20 *committees* o iteraciones para la obtención de resultados de todas las propiedades en ambas situaciones.

Cuadro 2. Predicción de pH, contenido de arcilla y cationes de intercambio con mínimos cuadrados parciales (PLSR), modelo cubista (CM), espectros cortados y transformados.

Parámetro	Set de datos	PLSR ¹			CM ¹		
		S	R ²	RMSE	S	R ²	RMSE
pH (-)	Calibración	0,07	0,26	0,45	0,04	0,36	0,42
	Validación	0	0,17	0,48	0,12	0,04	0,51
Contenido de arcilla (%)	Calibración	-2,65	0,08	8,63	0,16	0,05	8,77
	Validación	0	-0,01	8,02	-2,90	-0,01	8,04
Ca ⁺² intercambiabile (cmol ₍₊₎ kg ⁻¹)	Calibración	-0,54	0,50	6,73	-0,27	0,60	6,01
	Validación	0	0,51	6,82	-0,82	0,58	6,35
Mg ⁺² intercambiabile (cmol ₍₊₎ kg ⁻¹)	Calibración	-0,39	0,28	1,05	-0,01	0,46	0,91
	Validación	0	0,26	1,23	-0,39	0,29	1,21
Na ⁺ intercambiabile (cmol ₍₊₎ kg ⁻¹)	Calibración	-0,51	0,11	2,39	-0,25	0,66	1,47
	Validación	0	0,07	4,29	-0,38	-0,46	5,36
K ⁺ intercambiabile (cmol ₍₊₎ kg ⁻¹)	Calibración	0,02	0,40	0,41	-0,06	0,53	0,36
	Validación	0	0,16	0,40	-0,06	0,22	0,39

¹ S = sesgo; R² = coeficiente de determinación; RMSE = raíz cuadrada media del error.

Se observó, cuando se utilizaron los espectros sin transformación (como datos de reflectancia ocupando todo el espectro que abarca el equipo), que se obtuvieron mejores indicadores de R² y RMSE para todas las propiedades de suelo, a diferencia del uso de PLSR (Cuadro 3). En el modelo cubista, coincidentemente, los indicadores son mejores cuando se usan los espectros crudos, salvo en la estimación de Na intercambiabile (cmol₍₊₎ kg⁻¹) donde una transformación previa permitió disminuir el RMSE_{cal} en un 56,5%, pero con un aumento del RMSE_{val} del 15,7% con relación al espectro crudo. Cabe destacar también un aumento en los valores de R² para ambos conjuntos de datos (calibración y validación) para sodio intercambiabile. Una vez que se realizó el tratamiento; se mejoró este indicador desde valores que no representaban ajuste a 0,66 y 0,46 para R²_{cal} y R²_{val}, respectivamente.

En cuanto a los dos modelos utilizados en este estudio, el uso de CM generó que, cuando los espectros son tratados, se evidencia una disminución de la RMSE en las estimaciones de

Ca^{+2} , Mg^{+2} y K^{+} intercambiables, en ambos conjuntos de datos, donde las disminuciones en este parámetro estuvieron en los rangos entre el 7 al 13%. En cambio, cuando los espectros usados son crudos, solo la estimación de Mg intercambiable presentó una disminución consistente en ambos conjuntos de datos, con relación al otro modelo (PLSR).

El modelo PLSR ha sido usado ampliamente en la espectrometría (Minasny *et al.*, 2011; Sepúlveda *et al.*, 2021). Este consiste en un modelo lineal que predice propiedades de suelo mediante la extracción del espectro de factores ortogonales, variables latentes o componentes principales (Wadoux *et al.*, 2021). El modelo cubista, en tanto, tiene entre sus ventajas una interpretación simple, ya que selecciona cada longitud de onda para realizar sus predicciones, su funcionamiento es por medio de un algoritmo de árbol de regresiones. Últimamente ha sido considerado como un modelo ideal para el análisis de datos de suelo (Minasny y McBratney 2008; Tang *et al.*, 2020).

Independiente del modelo utilizado, el uso de validación cruzada está ampliamente difundido en la literatura para mejorar los grados de ajuste de los modelos realizados al usar todas las muestras en la construcción del modelo (Tahmasbian *et al.* 2017; Malmir *et al.*, 2019 ; Mammadov *et al.*, 2020). Contrario a lo realizado en este estudio, los datos no son separados en dos grupos como validación y calibración, si no que todas las muestras son separadas en $K-1$ submuestras que son usadas en el modelo y su uso se sugiere cuando la cantidad de muestras es acotada (Wadoux *et al.*, 2021). En este estudio, a pesar de tener un número de muestras más pequeño, se privilegió separar las muestras en dos grupos, de manera de contar con una aproximación real cuando el equipo fuese usado en condiciones de terreno.

Las estimaciones de pH y arcilla obtenidas en este estudio fueron más bajas que aquellas reportadas en otros estudios (Tang *et al.*, 2020; Mammadov *et al.*, 2020; Lobos *et al.*, 2021; Sepúlveda *et al.*, 2021). Ng *et al.* (2020) utilizaron un total de 1.601 muestras para efectuar la calibración, muy por encima de los usados en este estudio. Sin embargo, esas muestras presentaban una alta variación en relación con las utilizadas en este estudio, donde en el caso de pH y contenido de arcilla los CV fueron de 46,8 y 18,9% comparados con 34,00 y 8,53% obtenidos para las mismas propiedades. Condiciones similares utilizaron Mammadov *et al.* (2020) para calibrar, donde tanto el contenido de arcilla y el pH presentaban una alta variabilidad similares a los usados en este estudio (35,0 y 6,3), pero con un total de 114 muestras usando además validación cruzada. En ambos casos las calibraciones tuvieron mejores indicadores con R^2 de 0,83 y 0,54 (Mammadov *et al.*, 2020) y R^2 0,52 y 0,60 (Ng *et al.*, 2020) para contenido de arcilla y pH, respectivamente. El uso de rango desde 500-2.500 nm para estimación de pH, P y CIC tienen bajas estimaciones en comparaciones con COS (Stenberg *et al.*, 2010; Wijerwardane *et al.*, 2016). A pesar de valores bajos de R^2 alcanzados, un RMSE de 0,31 parece ser suficiente para un modelo de Vis-NIR, donde este error puede ser reducido usando calibraciones sitio-específicas (Soriano-Disla *et al.*, 2014).

Cuadro 3. Predicción de pH, contenido de arcilla y cationes de intercambio con mínimos cuadrados parciales (PLSR) y modelo cubista (CM) usando espectros sin transformación.

Parámetro	Set de datos	PLSR ¹			CM ¹		
		S	R ²	RMSE	S	R ²	RMSE
pH (-)	Calibración	0	0,35	0,43	0,02	0,33	0,43
	Validación	0,03	0,34	0,43	0,11	0,17	0,49
Contenido de arcilla (%)	Calibración	-2,36	0,17	8,19	0,12	0,37	7,24
	Validación	0	0,25	7,31	-2,15	0,09	8,39
Ca ⁺² intercambiable (cmol ₍₊₎ kg ⁻¹)	Calibración	-0,54	0,61	5,91	-0,13	0,64	5,79
	Validación	0	0,60	6,20	-1,52	0,61	6,31
Mg ⁺² intercambiable (cmol ₍₊₎ kg ⁻¹)	Calibración	-0,50	0,33	1,01	0,08	0,36	0,98
	Validación	0	0,38	1,25	-0,41	0,39	1,20
Na ⁺ intercambiable (cmol ₍₊₎ kg ⁻¹)	Calibración	-0,66	0,12	2,37	-0,64	NA	2,60
	Validación	0	0,10	4,28	-1,32	NA	4,63
K ⁺ intercambiable (cmol ₍₊₎ kg ⁻¹)	Calibración	-0,08	0,40	0,41	-0,04	0,70	0,32
	Validación	0	0,23	0,23	-0,06	0,28	0,38

¹ S = sesgo; R² = coeficiente de determinación; RMSE = raíz cuadrada media del error.

En el caso de los cationes intercambiables, las predicciones sí fueron más cercanas a lo observado en la literatura (Soriano-Disla *et al.*, 2013; Tang *et al.*, 2020; Malmir *et al.*, 2019; Mammadov *et al.*, 2020; Ng *et al.*, 2020). Ng *et al.* (2020) encontraron valores en el set de validación de 0,71, 0,59, 0,19 y 0,24 para Ca⁺², Mg⁺², K⁺ y Na⁺ intercambiables, respectivamente, utilizando modelo cubista. Hallazgos similares obtuvieron Tang *et al.* (2020) al evaluar estas propiedades en 400 muestras en Australia, usando también modelo cubista para los mismos cationes con valores de 0,63, 0,70, 0,29 y 0,45, respectivamente.

Una menor exactitud en la estimación de las arcillas podría estar asociada al material parental del suelo, ya que cuando la tefra volcánica contiene diferentes grados de meteorización de los minerales, estos no son activos a los rangos de medición de 1.300-2.600 nm (Fiantis *et al.*, 2019). Sin embargo, este estudio no considera suelos con estas características, salvo aquellos ubicados en la localidad de Cumpeo, pero como se mencionó anteriormente el mineral predominante sería la haloisita. Otro factor para considerar es el rango de medición, donde se ha visto que mediciones hechas el infrarrojo medio (MIR) tendría una mejor estimación comparado con una hecha en el infrarrojo cercano (NIR) (Soriano-Disla *et al.*, 2014).

Otro factor que explicaría los resultados es que el rango utilizado para la estimación de cationes de intercambio sea sub-óptimo (Pinheiro *et al.*, 2017; Lobos *et al.*, 2021). Sin embargo, estimaciones llevadas a cabo por Bikindou *et al.* (2014) en un rango de 1.100-2.500 nm obtuvieron estimaciones de K⁺ intercambiable con un R² de 0,78. En este estudio la estimación de algunas propiedades se realizó con longitudes de onda cercanas a los 2.500 nm (Cuadro 4), por lo tanto, se respalda el uso de los espectros completos para la calibración. Conjuntamente se ha determinado que una estimación de cationes intercambiables en el infrarrojo cercano ha presentado un R² promedio de 0,72, pero con algunas particularidades

por elemento (Soriano-Disla *et al.*, 2014). Lo que en general se observa es que en el caso de Ca y Mg intercambiable las estimaciones son mejores (en ese orden) que las del K y Na al igual que en este estudio (Soriano-Disla *et al.*, 2014; Tang *et al.*, 2020; Ng *et al.*, 2020).

Cuadro 4. Longitudes de onda usadas en los modelos de mínimos cuadrados parciales (PLSR) y modelo cubista (CM) con espectro completo.

Parámetro	PLSR	CM
pH (-)	2500;2000;1900;1400	2200; 1600
Contenido de arcilla (%)	2500;2200;1900	2200;1630;1500
Ca ⁺² intercambiable (cmol ₍₊₎ kg ⁻¹)	2500;2200;1900;1343	2500;1950;1400
Mg ⁺² intercambiable (cmol ₍₊₎ kg ⁻¹)	2400;2150;1900;1400	2400;2100;1600
Na ⁺ intercambiable (cmol ₍₊₎ kg ⁻¹)	2200;1900;1400	-*
K ⁺ intercambiable (cmol ₍₊₎ kg ⁻¹)	2500;2200;1900;1343	2400;2200

*: no utilizado

El Ca y el Mg podrían ser bien estimados en el rango del infrarrojo cercano, mientras que una mejor estimación de K se obtendría incluyendo el espectro visible (Vis-NIR), donde la relación del K con infrarrojo cercano se explicaría debido a la absorbancia de las ilitas (Soriano-Disla *et al.*, 2014). La estimación de K intercambiable está influenciada por factores tales como el K⁺ en solución que no son activos en el infrarrojo. Lo anterior queda de manifiesto ya que, al realizar una fertilización potásica, este elemento es más susceptible de sufrir pérdidas por procesos de lixiviación que Ca o Mg (Soriano-Disla *et al.*, 2014). Adicionalmente, se debe considerar que al tener una menor concentración en relación con Ca⁺² y Mg⁺², el efecto del K⁺ puede ser ocultado o enmascarado por estos cationes en la determinación (Minasny *et al.*, 2009).

El caso de Na⁺ es algo distinto, ya que este catión junto con otros tales como Cu, Fe, Mn y Zn poseen en general malas estimaciones en sus fracciones intercambiables, ya que se encuentran preferentemente en solución más que en la matriz del suelo (Janik *et al.*, 1998). Buenas estimaciones se han obtenido de Na⁺ intercambiable en el rango de rojo cercano (NIR) con un R² de 0,64, similares a los encontrados en este estudio cuando se procesó el espectro (Cuadro 2). Las estimaciones de Na permiten constatar que su calibración es específica al tipo de suelo y al medio ambiente (Minasny *et al.*, 2009).

Uno de los factores que adicionalmente pudiesen estar afectando los resultados, es el procesamiento de datos. En esta línea, Malmir *et al.* (2019) encontraron que el grado de ajuste del modelo obtenido mediante PLSR es mejor, cuando se comparó el uso de SNV con un espectro sin procesar. Dependiendo la propiedad de suelo que se quiere estimar, se pueden ir probando algunas transformaciones para mejorar su estimación. Por ejemplo, la estimación de los cationes K⁺, Mg⁺², Ca⁺² y Na⁺ fueron mejores con una transformación de corrección señal ortogonal (OSC) (Malmir *et al.*, 2019). Tahmasbian *et al.* (2017) obtuvieron mayores valores de R² al usar OSC para estimar N total y COS. Por contraparte, el uso de SNV entregó una mejor predicción de B, Cu y Zn (Malmir *et al.*, 2019), como para COS y contenido de

arcilla (Minasny *et al.*, 2011). Las diferencias en el ajuste del método a utilizar se pueden atribuir a variaciones en la reflectancia de los elementos y la distribución de sus correspondientes datos de referencia (Malmir *et al.*, 2019).

Un rango de 1.200-2.500 nm para el uso de NeoSpectra aparece como un buen predictor para propiedades de suelo, esto es porque muestra patrones similares de longitudes de onda con absorciones altas en los rangos de 1.400, 1.900 y 2.200 nm al igual que otros equipos. Tang *et al.* (2020) observaron que al usar un equipo NeoSpectra su predicción de contenido de arcilla podía ser menor por poseer un rango más acotado de 1250 nm-2500 nm, aun cuando los tres peak estaban dentro de la medición. En este estudio, el rango utilizado (1.442 nm - 2.284 nm) fue aún más acotado, lo cual explicaría en parte el bajo valor de R^2 alcanzado.

Otro punto es la resolución de medición del equipo, donde en la mayoría de la literatura se habla de rangos que van desde los 1-10 nm entre mediciones (Pinheiro *et al.*, 2017; Malmir *et al.*, 2019; Mammadov *et al.*, 2020; Ng *et al.*, 2020; Tang *et al.*, 2020). En el estudio que se presenta, la resolución del equipo fue de 24 nm, lo cual podría entregar una información que si bien es útil para hacer algunas estimaciones como cationes intercambiables, no sería suficiente para lograr altos valores de predictibilidad en arcilla y pH como se ha discutido anteriormente.

En general, en este estudio se ha visto que no existen absorciones significativas observables bajo los 1.400 nm y sobre los 2.300 nm. Dado que las longitudes de onda asociadas a la mineralogía de arcillas se centran entre los 1.400 y 2.200 nm, es posible correlacionarlo con la estimación de otras propiedades de suelo y ver su eficacia (Cohen *et al.*, 2007) (Cuadro 4).

Creación de matriz de corrección de humedad (EPO)

De acuerdo con la Figura 7a, los rangos de contenido de agua de suelo alcanzados en este estudio han sido ya usados para efectuar el procedimiento EPO (alrededor del 50% $g\ g^{-1}$) (Liu *et al.*, 2020; Wang *et al.*, 2020), logrando un buen ajuste al igual que con el uso de rangos de contenido de agua más estrechos (Minasny *et al.*, 2011). En general, se observa que a medida que el suelo se va secando, la absorbancia va disminuyendo. Cuando este fenómeno es analizado desde el parámetro de la reflectancia, el fenómeno ocurre a la inversa, esto es, el agua induce un aplanamiento del espectro a medida que esta va aumentando. Similares hallazgos fueron hechos por Chakraborty *et al.* (2019) al evaluar diferentes niveles de contenido de agua en suelos de Hungría.

Este fenómeno puede explicarse, ya que el agua se adsorbe como una capa delgada en la superficie a su vez que llena los poros, cambiando el índice de refracción del medio alrededor de las partículas del suelo desde el aire al agua. El bajo contraste resultante causa más dispersión hacia adelante y el largo del camino para reemerger el rayo emitido se vuelve más largo, aumentando la absorbancia (Twomey *et al.*, 1986).

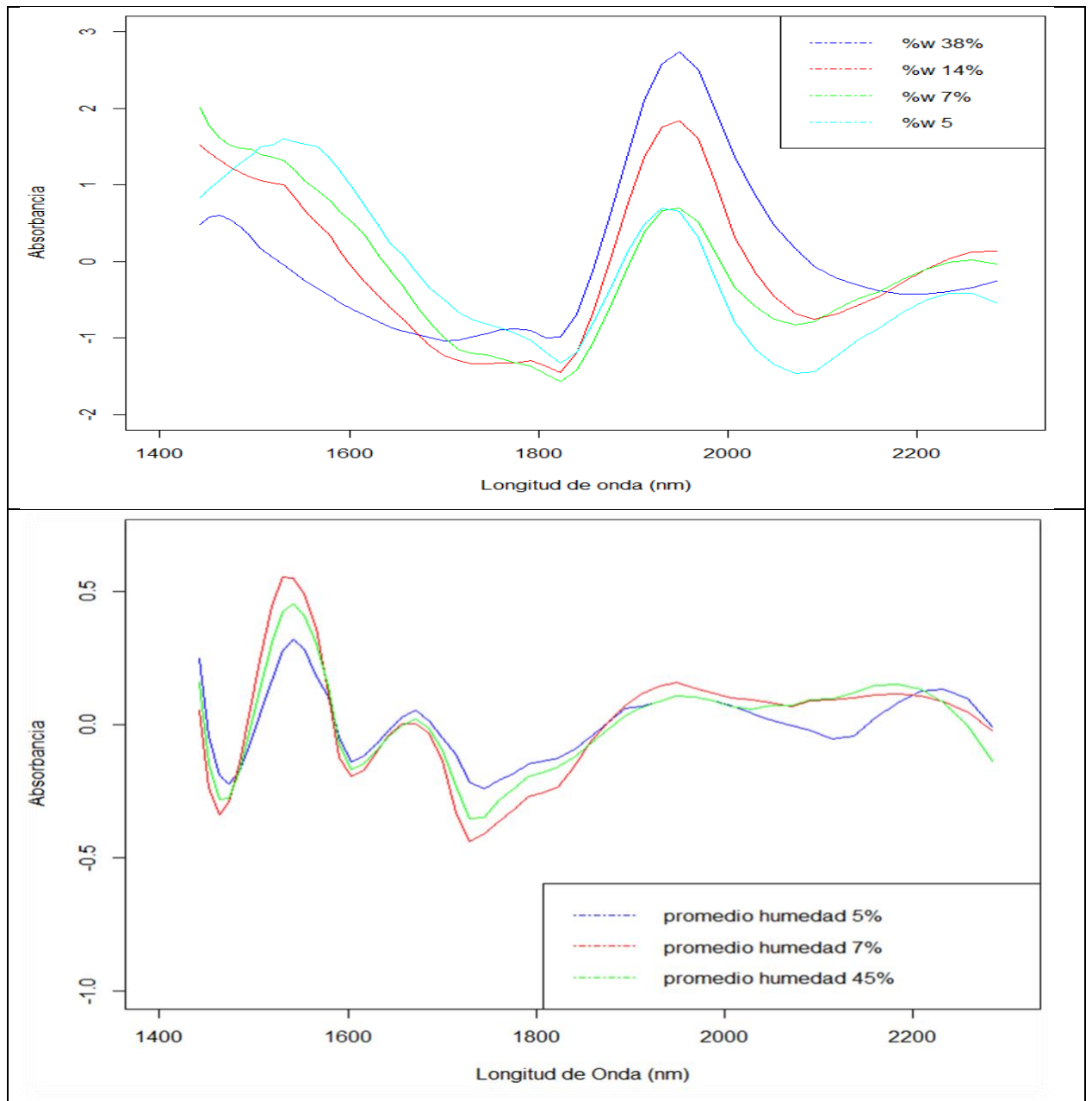


Figura 7. Absorbancia de los espectros 1, 2, 3 y 6 días luego de la hidratación para construir el parámetro de ortogonalización externo (EPO) (arriba)[a] y absorbancia de los espectros 1, 2, 3 y 6 luego de aplicar EPO sobre los mismos (abajo)[b] para el espectro 15.

Los rangos donde el valor de absorbancia es más alto coinciden con aquellos del rango 1.400 y 1.900 nm, destacándose un aumento leve alrededor de 1.780 nm. Estas bandas coinciden a su vez, con aquellas que son más sensibles en la determinación de propiedades de suelos en los estudios de espectrometría como en este estudio (Figura 8) (Soriano-Disla *et al.*, 2014). Wang *et al.* (2020) encontraron similares resultados al evaluar un equipo con rangos de medición más amplios, donde adicionalmente se observó un comportamiento similar alrededor de los 970 nm y 1.200 nm. Estos rangos se encuentran asociados a una fuerte adsorción de agua, lo que se explica por las tres frecuencias de vibración fundamentales del

agua: simétricos y asimétricos por la extensión del enlace O-H, como también la flexión de este enlace (Seema *et al.*, 2020; Wang *et al.*, 2020).

Otro factor que está afectando el comportamiento de una muestra o entre ellas es el contenido de MOS, el cual disminuye las capas de agua que rodean las partículas de suelo (Knadel *et al.*, 2014). Este fenómeno se explica ya que al tener un suelo un mayor contenido de MOS posee una mayor superficie específica, lo cual permite una menor cantidad de capas de agua reduciendo la profundidad a la cual la luz penetra la muestra o el suelo. Esto decrece la absorbancia por el agua, minerales de arcilla, y MOS en la región del infrarrojo cercano (Wang *et al.*, 2020).

En Figura 7b se puede observar que los espectros de suelo luego de la corrección de EPO mostraron tendencias similares a pesar de tener niveles de contenido de agua distintos. Así, los peak observados alrededor de los 1.400 y 1.900 nm ya no son visibles o se ven suavizados. Resultados similares se han observado con un número de muestras sometidas a rehidratación mayor (120-150) para la determinación de arcilla y COS (Minasny *et al.*, 2011, Chakraborty *et al.*, 2019).

Para la construcción de la matriz de corrección o proyección EPO se usaron los primeros 5 componentes principales para realizar el estudio (Cuadro 5). La decisión de cuántos componentes se utilizaron en la creación de la matriz de corrección, se determinó mediante el grado de ajuste que tienen las curvas con diferentes grados de contenido de agua luego de ponderar los espectros con la matriz creada (Figura 7a y 7b). Adicionalmente, se calculó y graficó el valor de Wilks, el cual permite aseverar que desde el uso de 5 componentes principales el valor se acerca a 1.

La Figura 8 permite saber cuáles son las bandas que tuvieron más influencia por el agua o condiciones externa. Las áreas con altos valores positivos y negativos son aquellas que deben ser transformadas (Minsany *et al.*, 2011; Wadoux *et al.*, 2021). El espectro transformado sin estas características es llamado XP. Una vez transformado, este puede ser usado en la modelación.

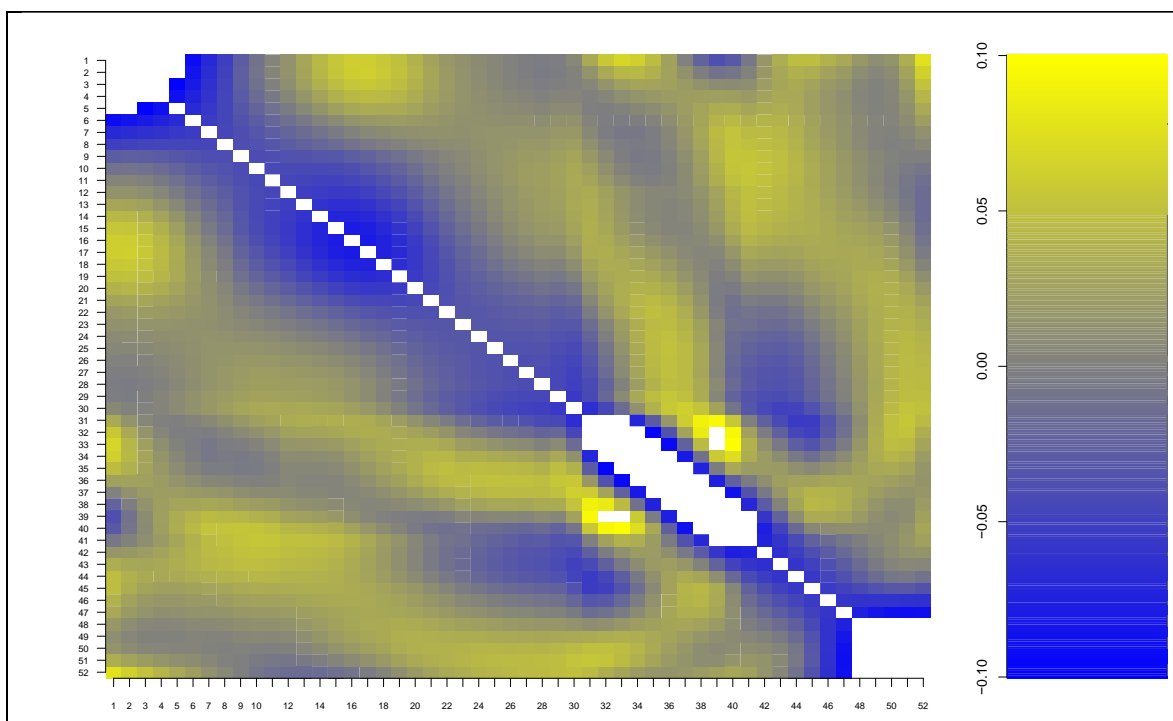


Figura 8. Proyección de la matriz P resultante del EPO. Los ejes son los índices de las ondas.

Cuadro 5. Estadígrafos de propiedades de suelos usado en la construcción de la matriz de proyección EPO.

Parámetro	Estadígrafos			
	X	Mín	Máx	CV
pH (-)	6,6±0,41	5,93	7,50	6,33
Contenido de arcilla (%)	27,3±9,40	11,70	50,60	34,40
Ca ⁺² intercambiable (cmol ₍₊₎ kg ⁻¹)	17,2±9,20	4,30	41,20	53,80
Mg ⁺² intercambiable (cmol ₍₊₎ kg ⁻¹)	2,4±1,50	0,08	6,61	61,90
Na ⁺ intercambiable (cmol ₍₊₎ kg ⁻¹)	3,0±4,60	0,87	29,88	151,00
K ⁺ intercambiable (cmol ₍₊₎ kg ⁻¹)	0,74±0,54	0	2,47	72,75

¹ X = promedio; ± : desviación estándar ; Mín = mínimo; Máx = máximo; CV = coeficiente de variación.

En el Cuadro 6, se puede observar que al aplicar EPO para la estimación de pH y contenido de arcilla el modelo cubista se logró disminuir el RMSE, al contrario de PLSR en los sets de validación. Por el contrario, en el caso de la estimación de cationes intercambiables el uso de una corrección de humedad generó un aumento en el RMSE en las estimaciones en el set de validación de Ca, Mg y K, pero no así en Na donde disminuyó para ambos modelos.

Se observaron diferencias en las estimaciones de Mg intercambiable al efectuar la calibración con PLSR con una disminución del 58,4% en el $RMSE_{cal}$ luego de aplicar una corrección por humedad, de igual manera que al calibrar de Ca intercambiable (-70% en el $RMSE_{cal}$), Na^+ intercambiable (-67,5%) y K intercambiable (-36,5%). Al aplicar el modelo cubista en tanto, se observó un aumento en el RMSE en la estimación de Mg^{+2} y Ca^{+2} intercambiables, mientras que en el resto de las propiedades se registró una disminución en este estadígrafo.

Para probar el efecto de la matriz de corrección, se utilizó el mismo conjunto de muestras secas que se usaron para efectuar el EPO. De esta manera, se efectuó la validación con las muestras con mayor contenido de agua, esto es luego de un día de la hidratación (EPO 1). A ambos conjuntos de datos no se les aplicó la corrección de humedad y fueron designados como “húmedo” en el Cuadro 6. Cuando se realizó el modelo mediante PLSR, el $RMSE_{val}$ entre las muestras húmedas (“Húmedo”) y aquellas que fueron corregidas por la matriz EPO (EPO val), la predictibilidad disminuyó para pH, contenido de arcilla y cationes de intercambio (Ca^{+2} , Mg^{+2} , K^+ , Na^+). En cambio, cuando se realizó la misma comparación, pero calibrando con el modelo cubista se registraron disminuciones de $RMSE_{val}$ en la estimación de pH (63,5%), contenido de arcilla (73,2%) y Mg^{+2} intercambiable (78,2%), y K intercambiable (13,7%), pero también aumentos en el RMSE como Ca^{+2} intercambiable (3,5%) y Na^+ intercambiable (21,8%).

En este sentido, el uso del radio de mejoramiento entre cuartiles (RPIQ) permitió inferir que dentro de los modelos efectuados mediante PLSR las estimaciones mediante el set de validación con EPO son todas insuficientes ($RPIQ < 1,4$), mientras que en los sets de calibración solo el modelo de Na^+ resulta ser insuficiente. Cabe mencionar que, en las muestras secas, salvo las estimaciones de contenido de arcilla y K en el set de validación, resultaron ser buenas estimaciones ($RPIQ > 1,4$). En el modelo cubista en tanto, los RPIQ mostraron un modelo suficiente con el set de validación solo con contenido de arcilla y Mg intercambiable, mientras que todas las propiedades mostraron modelos suficientes en el set de calibración (figuras 9 y 10).

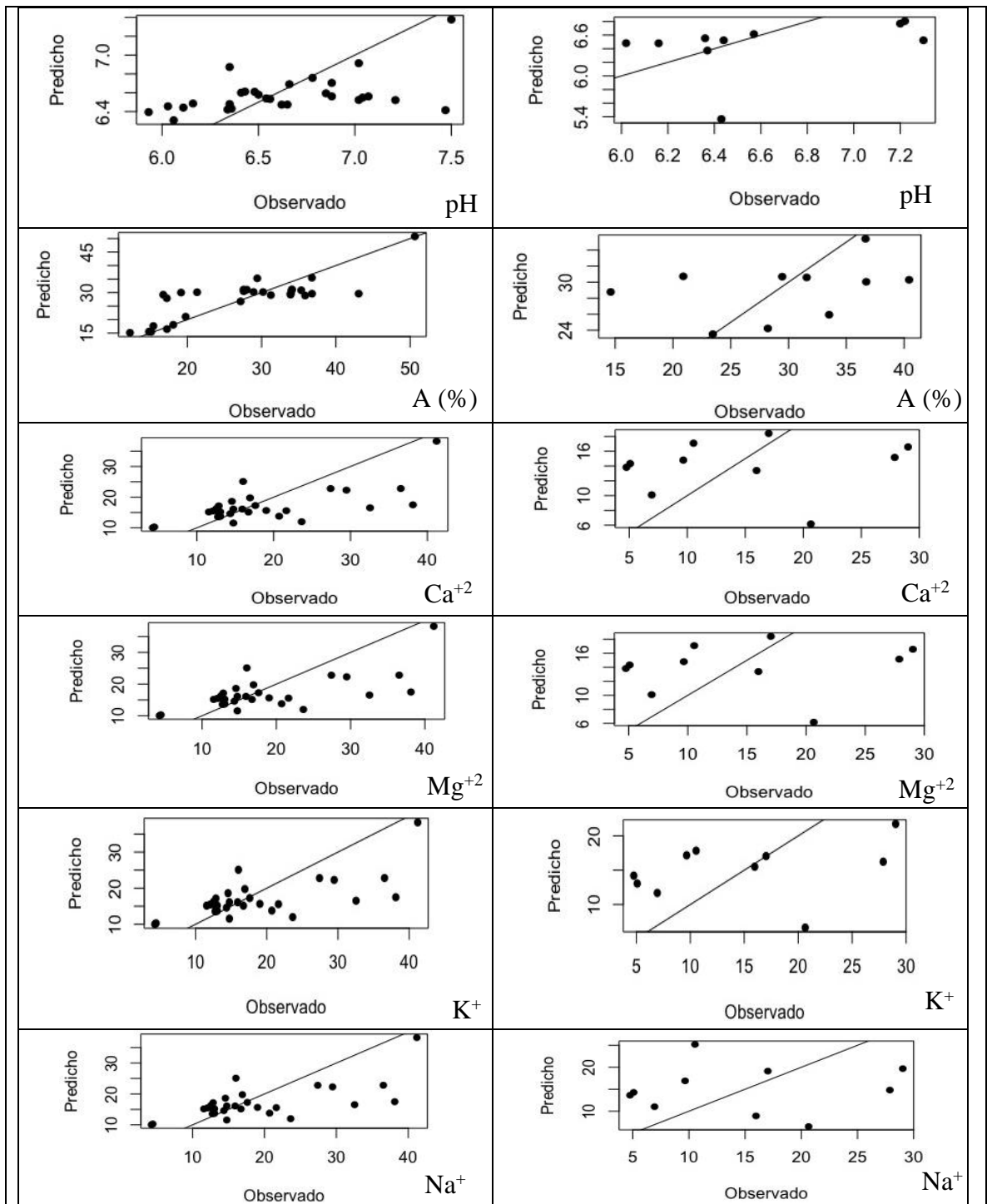


Figura 9. Predicción de pH, contenido de arcillas (A%) y cationes de intercambio (Ca^{+2} , Mg^{+2} , K^{+} y Na^{+}) en el set de calibración (lado izquierdo) y en el set de validación (lado derecho) luego de corregir con EPO mediante modelo cubista (CM).

Cuadro 6. Medidas de ajuste para predecir pH, arcillas y cationes de intercambio (Ca^{+2} , Mg^{+2} , K^+ y Na^+) con las muestras secas (validación) y después de efectuar el EPO mediante el modelo de mínimos cuadrados parciales (PLSR) y modelo cubista (CM). () indica cuál es el nivel de humedad que obtuvo la mejor estimación en los datos de calibración (Cal) y validación (Val).

Variable	Tratamiento	PLSR ¹			CM ¹		
		RPIQ	S	RMSE	RPIQ	S	RMSE
pH (-)	Secas Cal	1,48	0,03	0,43	1,45	0,02	0,43
	Secas Val	1,48	$1,7e^{-15}$	0,43	1,28	0,11	0,49
	Húmedo	0,19	2,40	2,47	0,50	1,29	1,37
	EPO Cal (2)	3,37	0	0,16	1,49	-0,04	0,36
	EPO Val (2)	1,25	0,02	0,54	1,37	-0,16	0,50
Arcilla (g 100g ⁻¹)	Secas Cal	1,11	-2,36	8,19	1,70	0,12	7,24
	Secas Val	1,11	$1,42e^{-14}$	7,31	1,53	-2,15	8,39
	Húmedo	0,21	38,40	39,05	0,42	25,89	26,96
	EPO Cal (1)	3,43	0	4,65	2,84	0,53	5,61
	EPO Val (1)	1,26	-2,99	8,93	1,55	-0,53	7,23
Ca intercambiable (cmol ₍₊₎ kg ⁻¹)	Secas Cal	1,78	-0,54	5,91	1,82	-0,13	5,79
	Secas Val	1,54	$1,7e^{-14}$	6,20	1,52	-1,52	6,31
	Húmedo	0,4	28,71	30,12	1,42	1,15	8,54
	EPO Cal (3)	5,01	0	1,76	1,25(5)	-1,63	7,05
	EPO Val (3)	1,07	2,20	11,33	1,37(5)	-0,77	8,84
Mg intercambiable (cmol ₍₊₎ kg ⁻¹)	Secas Cal	1,46	-0,50	1,01	1,49	0,07	0,98
	Secas Val	1,55	$6,2e^{-15}$	1,25	1,61	-0,41	1,20
	Húmedo	0,15	8,00	8,09	0,31	3,50	3,77
	EPO Cal (1)	4,03	0	0,42	1,57(5)	-0,04	1,07
	EPO Val (1)	0,44	-0,88	2,69	1,45(5)	-0,04	0,82
K intercambiable (cmol ₍₊₎ kg ⁻¹)	Secas Cal	1,81	-0,07	0,41	2,32	-0,04	0,32
	Secas Val	1,25	$2,1e^{-15}$	0,40	1,31	-0,06	0,38
	Húmedo	0,66	0,67	0,80	1,03	0,25	0,51
	EPO Cal (3)	1,99	0	0,26	2,51(2)	-0,01	0,20
	EPO Val (3)	0,69	0,50	0,76	1,21(2)	0,19	0,44
Na intercambiable (cmol ₍₊₎ kg ⁻¹)	Secas Cal	0,41	-0,66	2,37	0,38	-0,64	2,60
	Secas Val	0,39	$1,3e^{-15}$	4,28	0,36	-1,32	4,63
	Húmedo	0,25	-1,80	4,16	1,35	0,11	0,78
	EPO Cal (2)	1,43	0	0,77	2,46(3)	-0,01	0,45
	EPO Val (2)	0,52	1,12	2,01	1,10(3)	0,54	0,95

¹ RPIQ = radio de mejoramiento de distancia entre cuartiles; S = sesgo; RMSE = raíz cuadrada media del error.

Uno de los mayores desafíos en el uso de los equipos portátiles de espectrometría es el efecto del contenido de agua (Wang *et al.*, 2020). Esto limita el uso de los modelos generados solo en laboratorio, donde las muestras ya están secas y tamizadas, haciendo impracticable su uso en terreno donde el suelo se encuentra hidratado y no pulverizado (Ge *et al.*, 2014; Chakraborty *et al.*, 2019).

El uso de EPO para eliminar el efecto de la contenido de agua de la muestra se ha utilizado con éxito para la estimación de propiedades de suelo (Minsany *et al.*, 2011). Cuando no se toma toda la variabilidad de las muestras para la creación de la matriz de corrección, las

predicciones resultan ser pobres o en otras palabras el $RMSE_{val}$ no disminuye lo suficiente. Al respecto Chakaborty *et al.* (2019) al comparar dos sets para efectuar EPO encontraron que aquel que representaba todo el espectro de variabilidad de sus muestras obtenía mejores calibraciones.

En este estudio se logró capturar la mayor variabilidad posible, lo cual queda de manifiesto al comparar los estadígrafos de las todas las muestras (Cuadro 1) y de las 39 que se utilizaron para la construcción de la matriz de corrección (Cuadro 5).

Los resultados demuestran una disminución significativa de los valores $RMSE_{val}$ entre las muestras que estaban húmedas antes y después de la corrección de EPO. Sin embargo, los grados de ajuste luego de aplicarlo, por lo general nunca llegan a ser exactamente iguales a aquellas muestras que estaban secas (Ackerson *et al.*, 2017). Similares hallazgos han sido reportados por Ackerson *et al.* (2015), donde luego de efectuar la matriz de corrección tuvieron una disminución del 71% del RMSE y un 63% en el sesgo en suelos tropicales de Brasil. Chakaborty *et al.* (2019) estimaron una caída del 29% del RMSE cuando usaron una matriz de corrección que no abarcaba toda la variabilidad de sus muestras.

En general, la mayoría de las correcciones usadas se han hecho para las estimaciones de COS y contenido de arcilla mediante PLSR (Minasny *et al.*, 2011; Ge *et al.*, 2014; Ackerson *et al.*, 2017). En el caso de la estimación de contenido de arcilla y pH, el uso de EPO ha permitido buenos grados de ajuste. Sin embargo, en este estudio a pesar de la disminución del error por humedad, la estimación sigue siendo pobre en parte porque al observar las muestras en secas estas también tienen una baja predictibilidad (Cuadro 5). Lo anterior se explica porque EPO es un proceso de remoción de información espectral asociada a humedad; si la estimación en seco es pobre, esto significa una correlación pobre entre cierta propiedad y el espectro, por tanto EPO no tendrá una correlación fuerte por no haber información espectral de buena calidad (Wijerwardane *et al.*, 2016).

En cuanto al efecto de usar EPO en la corrección de estimación de cationes de intercambio, solo se puede tener una aproximación a través de la estimación de la CIC (Stenberg *et al.*, 2010; Wijerwardane *et al.*, 2016). Desafortunadamente, en estos casos, las estimaciones han sido pobres utilizando un rango de trabajo similar (500-2500 nm), y donde ocurre lo mismo que en el caso anterior: cuando la muestra está seca la estimación es pobre, una vez corregida lo seguirá siendo.

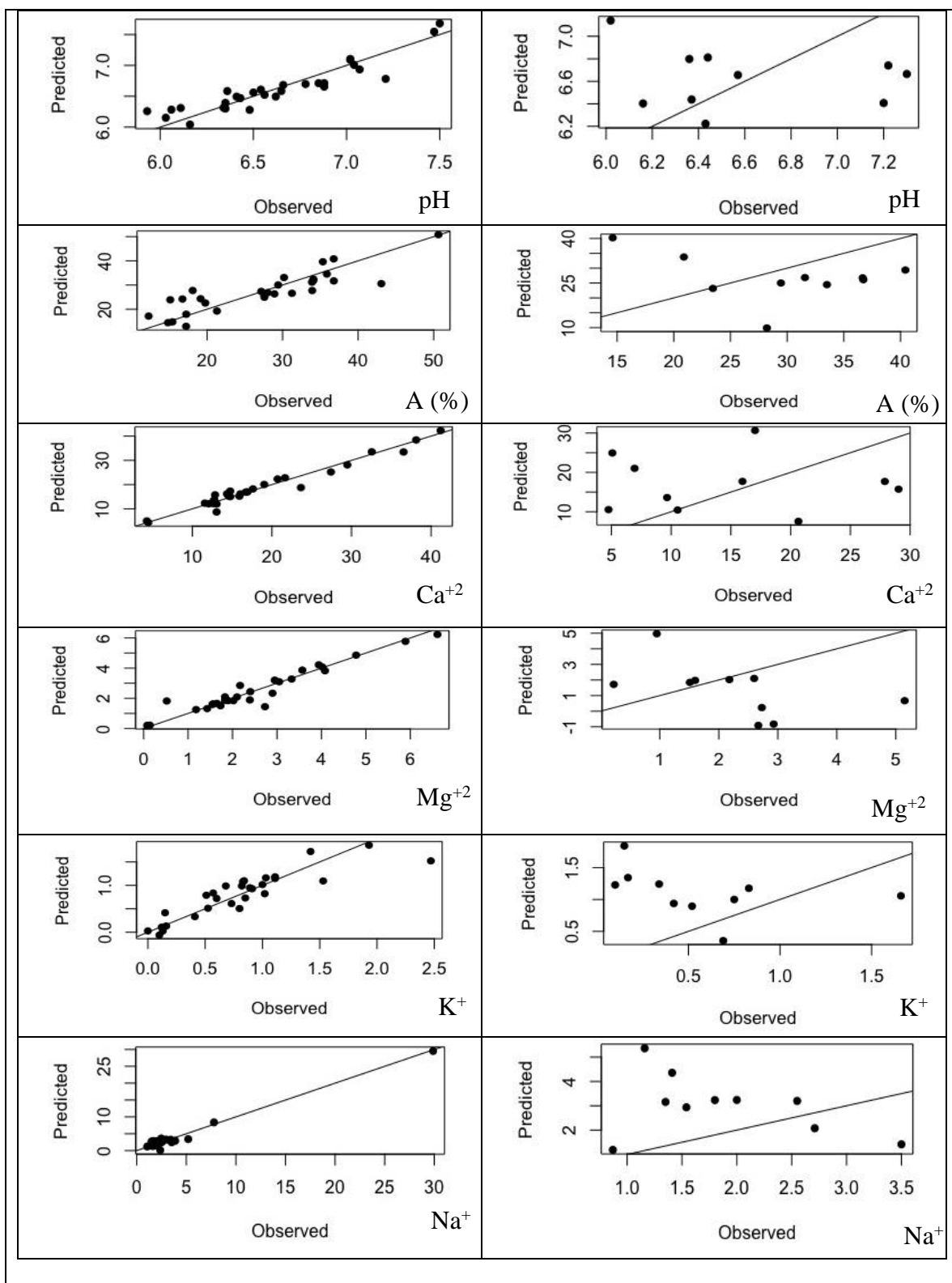


Figura 10. Predicción de pH, contenido de arcillas (A%) y cationes de intercambio (Ca^{+2} , Mg^{+2} , K^{+} y Na^{+}) en el set de calibración (lado izquierdo) y en el set de validación (lado derecho) luego de corregir con EPO mediante modelo de mínimos cuadrados parciales (PLSR).

La cantidad de muestras usadas para crear la matriz de corrección fueron bajas en relación con lo que se reporta en la literatura. Minasny *et al.* (2011) y Wijerwadane *et al.* (2016) usaron un total de 100 muestras las cuales fueron rehidratadas, por su parte Chakraborty *et al.* (2019) utilizaron un total de 150 donde algunas son rehidratadas y otras son usadas con la humedad de campo. Otra de las particularidades es que en estos tres estudios se utilizó un set de validación independiente al cual se usó para la creación de EPO. En este estudio, por la cantidad acotada de muestras utilizada se optó por usar las mismas muestras para realizar la calibración. Esto podría explicar, en parte, la alta disminución en el RMSE observados en el set de validación. Sin embargo, no se debe dejar de lado las otras variables que se consideraron en la creación de EPO discutidas anteriormente, como la representatividad de las muestras.

El uso de RPIQ permite ser más objetivo que al usar RMSE y el radio de mejoramiento de la desviación (RPD), ya que toma el error de predicción y la variación de los valores observados en cuenta a diferencia que los otros (Bellon-Maurel *et al.*, 2010; Wadoux *et al.*, 2010). Para el caso la espectrometría de suelos los valores debiesen ser superiores a 1,4. En este estudio los modelos presentaron valores por lo general por debajo de este límite y al observar las muestras secas los valores no superaron el límite para considerarlos como excelentes (>2) (Chang *et al.*, 2001). Esto confirma nuestros hallazgos anteriores y refuerza la idea de mejorar las calibraciones en seco.

A pesar de estos resultados, no deben dejar de ser tomados en cuenta por su aplicabilidad a su uso en terreno (Chakraborty *et al.*, 2019). La creación de una matriz EPO puede realizarse en terreno, ya que la agregación natural del suelo tiene una influencia limitada en los espectros relacionados a humedad (Ge *et al.*, 2014). Esto permitiría que se pudiese hacer con muestras intactas en terreno en conjunto con muestras rehidratadas facilitando labores logísticas (Chakraborty *et al.*, 2019).

CONCLUSIONES

La evaluación de un espectrómetro de infrarrojo cercano permitió obtener bajos niveles de predictibilidad de las propiedades de suelo tales como pH, contenido de arcilla y cationes intercambiables, respecto a los métodos de laboratorio con muestras secas al aire.

El uso de un parámetro de ortogonalización externo (EPO) permitió reducir el error experimental asociado al contenido de agua en el suelo en las estimaciones de pH, contenido de arcilla y cationes de intercambio por un espectrómetro de infrarrojo cercano portátil.

La predictibilidad de las propiedades de suelo tales como pH, contenido de arcilla y cationes de intercambio luego de corregir la contenido de humedad mediante un EPO es baja, al igual que cuando las muestras se encuentran secas.

El ajuste de algunos parámetros tales como cantidad de muestras usadas, el tipo de procesamiento de datos o el grado de resolución del equipo podrían mejorar la predictibilidad de estas variables en futuros estudios.

La correcta calibración de estos equipos permitiría que pudiesen ser usados para estimar propiedades de suelo de manera rápida y segura, permitiendo efectuar diagnósticos sitio-específicas.

LITERATURA CITADA

Ackerson, J.P., Demattê, J.A.M., Morgan, C.L.S. 2015. Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. *Geoderma* 259: 196–204.

Ackerson, J.P., Morgan, C.L.S., Ge, Y. 2017. Penetrometer-mounted VisNIR spectroscopy: application of EPO-PLS to in situ VisNIR spectra. *Geoderma* 286: 131–138.

Barthès, B.G., Brunet, D., Ferrer, H., Chotte, J.L., Feller, C. 2006. Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: influence of replication and sample grinding and drying. *Journal of Near Infrared Spectroscopy* 14 (5): 341–348.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A.B. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry* 29:1073–1081.

Bikindou, F., G., H., Philippe, D., Bouillet, J.P., Moukini, R., Mbedi, Y., Ngouaka, E., Brunet, D., Sita, S., Diazenza, J.B., Vouidibio, J., Mareschal, L., Ranger, J., Saint-Andre, L. 2012. Are NIR spectra useful for predicting site indices in sandy soils under Eucalyptus stands in Republic of Congo?. *Forest Ecology and Management* 266: 126–137.

Cardelli, V., Weindorf, D.C., Chakraborty, S., Li, B., De Feudis, M., Cocco, S., Agnelli, A., Choudhury, A., Ray, D.P., Corti, G. 2017. Non-saturated soil organic horizon characterization via advanced proximal sensors. *Geoderma* 288: 130-142

Casanova, M., Seguel, O., Salazar, O., Luzio, W. 2013. *Soils of Chile. Soils of the World Soils Serie.* Springer Science Business Media. The Netherlands. 183 p.

Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R. 2001. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Science Society of America Journal* 65: 480-490.

Chakraborty, S., Li, B., Weindorf, D. C., Morgan, C. L. S. 2019. External parameter orthogonalisation of Eastern European VisNIR-DRS soil spectra. *Geoderma* 337: 65–75.

Chang, C.W., You, C.F., Huang, C.Y., Lee, T.Q., 2005. Rapid determination of chemical and physical properties in marine sediments using a near-infrared reflectance spectroscopic technique. *Applied Geochemistry* 20: 1637–1647.

CIREN. 1996 . *Estudio Agrológico de la Región Metropolitana.* Centro de Información de Recursos Naturales. Publicación 115. Tomos I y II. 464 p.

CIREN. 1997. *Estudio Agrológico de la VII Región.* Centro de Información de Recursos

Naturales . Publicación 117. Tomos I y II. 586 p.

CIREN. 2002. Estudio Agrológico de la VI Región. Centro de Información de Recursos Naturales. Publicación 114. Tomos I y II. 546 p.

Cohen, M.J., Paris, J., Clark, M.W. 2007. P-sorption capacity estimation in Southeastern USA wetland soils using visible/near infrared (VNIR) reflectance spectroscopy. *Wetlands* 27:1098–1111.

Cooper, C.D., Mustard, J.F. 1999. Effects of very fine particle size on reflectance spectra of smectite and palagonitic soils. *Icarus* 142: 557-570.

Fiantis, D., Ginting, F.I., Gusnidar, Nelson, M., Minasny, B., 2019. Volcanic ash, insecurity for the people but securing fertile soil for the future. *Sustainability* 11(11): 3072.

Filzmoser, P., Garrett, R.G., Reimann, C., 2005. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences* 31: 579-587.

Gao, W., Yan, D. 2019. Warming suppresses microbial biomass but enhances N recycling. *Soil Biology and Biochemistry* 13: 111-118.

Ge, Y., Morgan, C.L.S., Ackerson, J.P. 2014. VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. *Geoderma* 221-222: 61-69.

Gmur, S., Vogt, D., Zabowski, D., Moskal, L.M. 2012. Hyperspectral analysis of soil nitrogen, carbon, carbonate, and organic matter using regression trees. *Sensors* 12(8): 10639–10658.

Gobrecht A., Bendoula, R., Roger, J-M., Bellon-Maurel, V. 2015. Combining linear polarization spectroscopy and the representative layer theory to measure the Beer-Lambert law absorbance of highly scattering materials. *Analytica Chimica Acta* 853:486–494.

Havlin, J., Tisdale, S., Nelson, W., Beaton, J. 2016. *Soil Fertility and Fertilizers: An introduction to nutrient management* (7th ed.). Prentice Hall, N.J. USA. 515 p.

Hirzel, J. 2014. Diagnóstico nutricional y principios de fertilización en frutales y vides. Colección libros INIA (Instituto de Investigaciones Agropecuarias) 31. 322 p.

Hirzel, J. 2011. Fertilización de cultivos en Chile. Colección libros INIA (Instituto de Investigaciones Agropecuarias) 28. 432 p.

Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T.F.A., McBratney, A.B., Pallasser, R., Pozza, L. 2015. Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. *Geoderma* 241–242: 180-209.

- Janik, L.J., Merry, R.H., Skjemstad, J.O. 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture* 38: 681–696.
- Knadel, M., Deng, F., Alinejadian, A., de Jonge, Lis Wollesen, Moldrup, P., Greve, M.H. 2014. The Effects of moisture conditions—from wet to hyper dry—on visible near-infrared spectra of Danish reference soils. *Soil Science Society of America Journal* 78(2): 422–433.
- Lobos, I., Alfaro, M., Martinez, J., Pavez, P. 2021. Caracterización química de suelos volcánicos utilizando espectroscopia de infrarrojo cercano (NIRS). *Chilean Journal Agricultural and Animal. Science. ex Agro-Ciencia* 37(1):32-42.
- Lin, Z.D., Wang, R.J., Wang, Y.B., Wang, L.S., Lu, C.P., Liu, Y., Zhang, Z.Y., Zhu, L.K., 2018. Accurate and rapid detection of soil and fertilizer properties based on visible/nearinfrared spectroscopy. *Applied Optics* 57(18): D69–D73.
- Liu, Y., Deng, C., Lu, Y., Shen, Q., Zhao, H., Tao, Y. 2020. Evaluating the characteristics of soil vis-NIR spectra after the removal of moisture effect using external parameter orthogonalization. *Geoderma* 376:114568.
- Luzio L., W., Casanova, M. y Seguel, Ó. 2010. *Suelos de Chile*. Santiago, Chile. 364 p. Disponible en <https://repositorio.uchile.cl/handle/2250/182305>
- Malmir, M., Tahmasbian, I., Xu, Z., Ferrar, M. B., Bai, S. 2019. Prediction of soil macro- and micro-elements in sieved and ground air dried soils using laboratory-based hyperspectral imaging techniche. *Geoderma* 340: 70-80.
- Mammadov, E., Denk, M., Riedel, F., Lewinska, K., Kaźmierowski, C., Glaesser, C. 2020 Visible and Near-Infrared Reflectance Spectroscopy for Assessment of Soil Properties in the Caucasus Mountains, Azerbaijan. *Communications in Soil Science and Plant Analysis* 51(16): 2111-2136.
- McBratney, A., de Gruijter, J., Bryce, A. 2019. Pedometrics timeline. *Geoderma* 338: 568-575.
- Minasny, B. andMcBratney,A.B. 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemometrics and. Intelligent. Laboratory. Systems* 94: 72–79.
- Minasny, B., McBratney, A.B., Pichon, L., Sun,W., Short, M.G. 2009. Evaluating near infrared spectroscopy for field prediction of soil properties. *Australian Journal of Soil Research* 47(7): 664–673.
- Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, Jean-Michael., Gobrecht, A., Ferrand, L., Joallanda, S. 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma* 167-168: 118-124.

Morgan, C.L.S., Waiser, T.H., Brown, D.J., Hallmark, C.T. 2009. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. *Geoderma* 151: 249-256.

Ng, W., Husnain., Anggria, L., Siregar, A.F., Hartatik, W., Sulaemna, Y., Jones, E., Minasny, B. 2020. Developing a soil spectral library using a low-cost NIR spectrometer for precision fertilization in Indonesia. *Geoderma Regional* 22: e00319.

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., Mouazen, A. M. 2016. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil and Tillage Research* 155: 510–522.

Pfeiffer, M., Padarian, J., Osorio, R., Bustamante N., Olmedo, G.F., Guevara, M., Aburto, F., *et al.* 2020. CHLSOC: the Chilean Soil Organic Carbon database, a multi-institutional collaborative effort. *Earth System Science Data* 12: 457–468.

Pinheiro, E.F.M., Ceddia, M.B., Clingensmith, C.M., Grunwald, S., Vasques, G.M. 2017. Prediction of soil physical and chemical properties by visible and near-infrared diffuse reflectance spectroscopy in the Central Amazon. *Remote Sensing* 9(4): 293.

Roger, J.M., Chauchard, F., Bellon-Maurel, V. 2003. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemometric and Intelligent Laboratory Systems* 66: 191-204.

Sadzawka, A., Grez, R., Carrasco, M.A., Mora, M. 2006. Métodos de análisis recomendados para los suelos de Chile. Revisión 2006. Instituto de investigaciones Agropecuarias, Series Actas INIA N°34, Santiago, Chile, 164 p.

Sadzawka, A. 2006a. Propiedades físico-químicas de los suelos II: Adsorción e Intercambio iónico. pp:134-168. In: Avances en el conocimiento de los suelos de Chile. (Re-edición, 2020). Luzio, W. y Casanova, M.(Eds.). Santiago, Chile. 460 p.

Sadzawka, A. 2006b. Química, diagnóstico y recuperación de suelos afectados por sales. pp: 169-206. In: Avances en el conocimiento de los suelos de Chile. (Re-edición, 2020). Luzio, W. y Casanova, M. (Eds.). Santiago, Chile. 460 p.

Sandoval, M., Dörner, J., Seguel, O., Cuevas, J., Rivera, D. 2012. Métodos de Análisis Físicos de Suelos. Universidad de Concepción, Publicaciones del Departamento de Suelos y Recursos Naturales N° 5. Chillán, Chile. 80 p.

Seema, Ghosh, A. K., Das, B. S., Reddy, N. 2020. Application of VIS-NIR spectroscopy for estimation of soil organic carbon using different spectral preprocessing techniques and multivariate methods in the middle Indo-Gangetic plains of India. *Geoderma Regional* 23: e00349.

Sepúlveda, M.A., Hidalgo, M., Araya, J., Casanova, M., Muñoz, C., Doetterl, S., Wasner, D., Colpaert, B., Bodé, S., Boeckx, P., Zagal, E. (2021). Near-infrared spectroscopy:

Alternative method for assessment of stable carbon isotopes in various soil profiles in Chile. *Geoderma Regional* 25: e00397.

Sharififar, A., Singh, K., Jones, E., Ginting, F.I., Minasny, B. 2019. Evaluating a low-cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models. *Soil Use and Management* 35: 607-616.

Siesler, H. W., Kawata, S., Heise, H. M., Ozaki, Y. 2008. Near-infrared spectroscopy: principles, instruments, applications. John Wiley & Sons. Germany-Weinheim. 347 p.

Soriano-Disla, J.M., Janik, L.J., Rossel, R.A.V., Macdonald, L.M., McLaughlin, M.J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews* 49(2): 139-186.

Stenberg, B., Rosel, R.A.V., Moauzen, A.M., Wetterlind, J. 2010. Chapter five-visible and near infrared spectroscopy in soil science. *Advances in Agronomy* 107: 163-215.

Stolpe N., Zagal E., Sandoval M., Quezada C. 2008. Southern Field-guide. The International Conference and Field Workshops on soil classification. TUSS-MINAGRI- U. de Chile- U. de Concepción- U. de La Serena-SChCS. (Casanova M., Luzio W., eds.). Santiago-Ovalle-Chillán 9-18 noviembre. 94 p.

Tahmasbian, I., Xu, Z., Boyd, S., Zhou, J., Esmaeilani, R., Che, R., Hosseini Bai, S. 2018. Laboratory-based hyperspectral image analysis for predicting soil carbon, nitrogen and their isotopic compositions. *Geoderma* 330: 254-263.

Tamburini, E., Vicenzi, F., Costa, S., Mantovi, P., Pedrini, P., Castaldelli, G. 2017. Effects of moisture and particule size on quantitative determination of total organic carbon (TOC) in soils using near-infrared spectroscopy. *Sensors* 17: 2366.

Tang, Y., Jones, E., Minasny, B., 2020. Evaluating low-cost portable near infrared sensors for rapid analysis of soils from South Eastern Australia. *Geoderma Regional* 20: e00240.

Twomey, S.A., Bohren, C.F., Mergenthaler, J.L., 1986. Reflectances and albedo differences between wet and dry surfaces. *Applied Optics* 25: 431-437.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C. 2011. Proximal soil sensing: an effective approach for soil measurements in space and time. *Advances in Agronomy* 113: 243-291.

Wadoux, A. M. J.-C., Malone, B., Minasny, B., Fajardo, M., McBratney, A. B. 2021. Soil Spectral Inference with R. *Progress in Soil Science*. Sydney, Australia. 257 p.

Wang, Y.-P., Lee, C.-K., Dai, Y.-H., Shen, Y. 2020. Effect of wetting on the determination of soil organic matter content using visible and near-infrared spectrometer. *Geoderma* 376: 114528.

Wang, D., Chakraborty, S., Weindorf, D.C., Li, B., Sharma, A., Paul, S., Ali, M.N. 2015. Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. *Geoderma* 243-244: 157-167.

Wijewardane, N.K., Ge, Y., Morgan, C. L.S. 2016. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. *Geoderma* 267: 92-101.

Xu, S., Zhao, Y., Wang, M., Shi, X. 2018. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis-NIR spectroscopy. *Geoderma* 310: 29-34.