Universidad de Concepción
Dirección de Postgrado
Facultad de Agronomía
Programa de Magíster en Ciencias Agronómicas

PROPIEDADES BIOQUÍMICAS DEL SUELO EN AGROECOSISTEMAS DE LA ZONA CENTRO SUR ESTIMADAS POR ESPECTROSCOPÍA VIS-NIR

Tesis para optar al grado de Magíster en Ciencias Agronómicas

JEAN IVÁN INTRIAGO ÁVILA
CHILLÁN-CHILE
2025

Profesor Guía: Erick Manuel Zagal Venegas, PhD
Dpto. de Suelos y Recursos Renovables
Facultad de Agronomía
Universidad de Concepción

Esta tesis ha sido realizada en el Departamento de Suelos y Recursos Naturales de la Facultad Agronomía, Universidad de concepción.

Profesor Guía

Erick Zagal Venegas, PhD
Facultad de Agronomía
Universidad de Concepción

Comisión Evaluadora:

Dr. Miguel Garriga Caraballo
Facultad de Agronomía
Universidad de Concepción

Dr. Lizardo Reyna Bowen
Facultad de Ingeniería Agrícola
Universidad Técnica de Manabí-Ecuador

Director de Programa

Macarena Gerding González, PhD
Facultad de Ciencias Agronomía
Universidad de Concepción

Dedicatoria

El presente trabajo se lo dedico, en primer lugar, a Dios, ya que gracias al he logrado una nueva etapa por enseñarme el camino correcto de la vida, guiándome y fortaleciéndome cada día.

A mis profesores guía, Dr. Erick Zagal, Dr. Miguel Garriga y Dr. Lizardo Reyna que me impartieron sus conocimientos, los cuales he ido asimilando para mi bienestar. Al Dr. José Reyna por la ayuda que me proporcionó.

A mi papá Iván Intriago porque buscó siempre soluciones para mi bienestar y apoyo condicional; a mi mamá Narcisa Ávila por estar siempre pendiente de mí e incentivarme a lograr más cosas; ambos creyeron en mí y me sacaron adelante, enseñándome ejemplos dignos de superación y entrega, porque en gran parte gracias a ustedes, hoy puedo ver alcanzada mi meta que siempre estuvieron impulsándome en los momentos más difíciles de mi Postgrado; a mi hermano Deytier Intriago por la gran ayuda que me proporcionó, va por ustedes, por lo que valen, porque admiro sus fortalezas y por lo que han hecho por mí en los momentos difíciles; a mi familia en especial a mi tío Felicísimo Intriago por apoyarme de una u otra manera y por último a mis amigos de Ecuador.

A la Familia Iglesias y en especial a Brenda Limongi Iglesias, estuvo en todos los momentos conmigo desde Pregrado hasta el Magíster y a pesar de la distancia

su amor permaneció intacto, su paciencia y su apoyo permanente.

Al equipo de trabajo que del laboratorio que lo conforman el profesor Erick, Marita, Katherine, Marcela, Camila y Carlos por su gran apoyo a este trabajo tanto en campo como en laboratorio.

Se agradece al Gobierno de Chile y al Proyecto FONDEF ID21I10264 el cual financió este proyecto de tesis de postgrado.

A la Facultad de Agronomía y al Programa de Magíster en Ciencias Agronómicas por aportar las enseñanzas adecuadas para la formación como profesionales en Ciencias.

A la Sra. Mirna Schafer y a los amigos/compañeros que he ido conociendo a lo largo del tiempo que han hecho mis días inolvidables en Chile.

Tabla de contenido

Índice de tablas, figuras y ecuaciones

Resumen

El uso de la espectroscopia del infrarrojo cercano visible (Vis-NIR) presenta un enfoque no destructivo, rápido, fiable y rentable para predecir las propiedades físicas, químicas y bioquímicas del suelo. En este estudio, se estimaron propiedades bioquímicas como la respiración basal del suelo, la biomasa microbiana activa (hidrólisis de diacetato de fluoresceína), la actividad β-glucosidasa y la materia orgánica particulada-C (POM), variables clave en el ciclo del carbono. Se muestrearon 70 sitios en la zona centro-sur de Chile, específicamente en las regiones del Maule, Biobío, Ñuble y Araucanía (35°58′ S 72°58′ O) en cultivos tradicionales, pastizales y frutales a una profundidad de 0-30 cm. Las muestras fueron analizadas en laboratorio y escaneadas usando un espectrofotómetro UV-Vis-NIR cubriendo un rango de 175 a 3300 nm y una resolución de 1 nm. Se utilizaron modelos multivariantes que incluían PLS, un método de regresión lineal, y Random Forest (RF), basado en el aprendizaje automático. Se evaluaron siete métodos de preprocesamiento, y la corrección ortogonal de la señal (OSC) resultó ser la mejor técnica de preprocesamiento en los modelos PLS. Los modelos PLS, con preprocesamiento OSC, y corrección multiplicativa de la dispersión (MSC), mostraron una alta precisión en rasgos como la respiración basal y la β-glucosidasa (R2validación: 0,99-0,98; RMSECV: 0,05-0,101, respectivamente). La FDA requirió transformaciones logarítmicas, mientras que la POM-C fue sensible a las correcciones espectrales (R^2validación: 0,95; RMSECV: 1,0486). Random Forest (RF) presentó resultados heterogéneos:

La POM-C se predijo bien (RPD = 2,08; R^2validación = 0,77, RMSEP:1,128), pero la biomasa y la β-glucosidasa tuvieron rendimientos bajos. Basándose en estos resultados, la espectroscopia del suelo muestra potencial como herramienta para la estimación de los atributos bioquímicos del suelo.

Abstract

The use of visible near-infrared spectroscopy (Vis-NIR) presents a non-destructive, fast, reliable, and cost-effective approach to predict soil physical, chemical and biochemical properties. In this study, biochemical properties such as basal soil respiration, active microbial biomass (Fluorescein diacetate hydrolysis), β-glucosidase activity and particulate organic matter-C (POM), key variables in the carbon cycle, were estimated. Seventy sites were sampled in south-central Chile, specifically in the regions of Maule, Biobío, Ñuble and Araucanía (35°58′ S 72°58′ W) in traditional crops, pastures, and fruit trees at a depth of 0-30 cm. The samples were analyzed in a laboratory and scanned using an UV-Vis-NIR spectrophotometer covering a range from 175 to 3300 nm and a resolution of 1 nm. Multivariate models including PLS, a linear regression method, and Random Forest (RF), which is based on machine learning, were used. Seven preprocessing methods were evaluated, and the orthogonal signal correction (OSC) emerged as the best preprocessing technique in PLS modelling. PLS models, with preprocessing OSC, and multiplicative scatter correction (MSC), showed high accuracy on traits such as basal respiration and β-glucosidase (R^2validation: 0.99-0.98; RMSECV: 0.05-0.101, respectively). The FDA required logarithmic transformations, while POM was sensitive to spectral corrections (R^2validation: 0.95; RMSECV: 1.0486). Random Forest (RF) presented heterogeneous results: POM was well predicted (RPD = 2.08; R2validation = 0.77, RMSEP:1.128), but biomass and β-glucosidase had low performances.

Based on these results, soil spectroscopy shows potential as a tool for the estimation of soil biochemical attributes.

Introducción general

El suelo es un material estratificado, presente en la superficie terrestre de la Tierra y es el resultado de procesos químicos, biológicos y de la organización física de los minerales y la materia orgánica (McBratney & Hartemink, 2024). Es un recurso natural finito, no renovable, que entrega innumerables servicios ecosistémicos como provisión de alimentos, hábitat de microrganismos, regulación de clima, entre otros (Burbano-Orjuela, 2016; Zabaloy, 2021). Sin embargo, la agricultura intensiva ha degradado los suelos y consecuentemente la salud y la calidad de estos.

La salud de los suelos está asociada a su capacidad continua para funcionar como un ecosistema, como base para el desarrollo de plantas, animales y humanos (NRCS, 2019). En general, un suelo sano incluye (i) labranza cero, (ii) suficiente profundidad de las raíces para acceder agua y nutrientes, (iii) pH óptimo para los suelos agrícolas y, (iv) una alta población y diversidad de microrganismos beneficiosos para la descomposición de la materia orgánica (MO) los ciclos de nutrientes y la conservación de la estructura del suelo (Magdoff, 2001; Mann et al., 2019).

La calidad del suelo en términos generales se define como la “capacidad de un suelo para preservar la productividad biológica, de un agroecosistema”. Esta

tiende a degradarse con el manejo intensivo, lo que implica una disminución de su capacidad para realizar una o más de sus funciones productivas o ecosistémicas, lo que resulta en una mayor fragilidad del suelo (Clunes *et al.,* 2022). Un componente central de la salud y calidad de los suelos es la materia orgánica (MO).

La materia orgánica, como fracción orgánica del suelo, influye en sus propiedades físicas, químicas y biológicas y constituye un reservorio de nutrientes, que por la acción de los microorganismos del suelo se descompone activamente entregando nutrientes para el crecimiento de las plantas (Walsh & McDonnell, 2012). La MO del suelo es un recurso dinámico y cambiante, que refleja el equilibrio entre la adición de nueva MO y la pérdida de la que ya está en el suelo, en parte controlada por la actividad biológica en el suelo (Cotching, 2018). El funcionamiento de la MO en el suelo es clave para la provisión de ciertos servicios ecosistémicos y un modelo conceptual reciente la categoriza en dos reservorios distintos con propiedades y comportamientos diferentes: materia orgánica particulada (POM, de su sigla en inglés) y materia orgánica asociada a la matriz mineral (MAOM, de su sigla en inglés) (Cotrufo & Lavallee, 2022). Brevemente, la POM se compone esencialmente de fragmentos ligeros que están relativamente descompuestos, mientras que la MAOM consiste en moléculas individuales o fragmentos microscópicos de material orgánico que se han lixiviado directamente del material vegetal o han sido transformados

bioquímicamente por la biota del suelo. La diferencia fundamental entre estos reservorios es que la MAOM está protegida de los procesos biológicos de descomposición a través de la asociación con los minerales del suelo, mientras que la POM no lo está. Las asociaciones minerales incluyen enlaces químicos entre las moléculas orgánicas precursoras y las superficies minerales y la oclusión al interior de microporos o pequeños agregados (<50-63 μm), que hacen que la MO del suelo sea menos accesible para los descomponedores y sus enzimas. Debido a esta diferencia fundamental en sus niveles de protección contra la descomposición, la MAOM tiende a persistir durante mucho más tiempo que la POM (Lavallee *et al.,* 2020).

El comportamiento de la biota, la actividad biológica, la fertilidad del suelo y el almacenamiento de carbono (C) son de gran interés en el escenario actual de cambio climático. La influencia del clima en la flora (Telo da Gama, 2023) y fauna del suelo (Sergeevna Kozun *et al.,* 2022), así como en la microbiota y la actividad enzimática (Kazeev *et al.,* 2020), son conocidos en diferentes suelos de zonas naturales, y las diferentes categorías taxonómicas tienen diferencias significativas en la diversidad de la biota del suelo, su composición y actividad biológica (Sergeevna Kozun *et al.,* 2022). Chile, presenta gran diversidad de suelos, teniendo 11 de los 12 órdenes de suelo descritos en la taxonomía de suelos del Departamento de Agricultura de los Estados Unidos (Staff, 1999). En Chile el centro-sur templado y lluvioso permite el desarrollo de Andisoles, suelos

derivados de cenizas volcánicas que tienen gran importancia, ya que representan el 50-60% de las tierras cultivables en el país (Valle & Carrasco, 2018). Por otro lado, suelos residuales y coluviales coexisten con suelos derivados de las cenizas volcánicas.

El carbono orgánico del suelo (COS) es uno de los mayores reservorios de C en los ecosistemas terrestres y su intercambio con la atmósfera puede variar según las condiciones climáticas y su interacción con propiedades geoquímicas del suelo, factores que actúan como control de la estabilización del C en la matriz mineral (Doetterl *et al.,* 2015). El COS lleva a cabo muchas funciones en los agroecosistemas, incluyendo actividades microbianas, es indicador de la salud del suelo, debido a sus importantes contribuciones a la calidad de los suelos, y contribuye a la mitigación al cambio climático. Bajo este contexto, es posible preguntarse si el suelo es una fuente o sumidero de dióxido de carbono (CO_2) atmosférico (Carvalhais *et al.,* 2014). Está bien documentado que los suelos biológicamente más activos se caracterizan por una mayor biomasa microbiana, lo que conduce a una mayor actividad enzimática del suelo y a una mayor respiración del suelo (Stark *et al.,* 2007), especialmente de la MO más lábil. Las enzimas del suelo permiten a los microorganismos acceder a la energía y los nutrientes presentes en sustancias complejas a través de procesos de degradación y mineralización, así como la formación de moléculas complejas y estables en procesos de humificación (MO estable). La mayoría de las

actividades enzimáticas del suelo son directamente proporcionales al contenido de COS nativo, lo que refleja comunidades microbianas mayores y la estabilización de enzimas en materiales húmicos. La actividad de la β-glucosidasa se utiliza con frecuencia como indicador de cambios en la cantidad y calidad del COS (Preethi *et al.,* 2013). Uno de los grandes desafíos que nos impone el cambio climático, es conocer con certeza y rapidez las propiedades del suelo, ya que se requiere monitoreo permanente de las propiedades químicas, físicas y biológicas del suelo para determinar su estado de salud para la actividad agrícola y la provisión de otros servicios ecosistémicos (Zhang *et al.,* 2023).

El suelo es un sistema heterogéneo donde los procesos que ocurren son difíciles y complejos de comprender en su totalidad (Viscarra Rossel *et al.,* 2006). Las mediciones tradicionales de laboratorio (química húmeda) son costosas, requieren mucho tiempo de mano de obra especializada y son propensas a producir residuos químicos y contaminación ambiental por tanto, se dificulta satisfacer las necesidades de un monitoreo de manera permanente y a gran escala para determinar el estado y de los suelos (Song *et al.,* 2024). Para poder generar información de manera más rápida y eficiente, diversas metodologías han sido objeto de interés. Una de las que ha tenido mayor desarrollo durante las últimas dos décadas, es la espectroscopía de suelos (química seca).

La espectroscopía de suelo es un potente método analítico que se basa en la interacción entre la materia y la radiación electromagnética a frecuencias específicas, lo que resulta en una firma espectral única (Shaw & Burke, 2003). El acople de tecnologías espectroscópicas con quimiometría (construcción de modelos predictivos) proporciona las herramientas necesarias para el análisis de datos y diagnóstico de la salud de los suelos (Barra *et al.,* 2021).

La espectroscopía visible-infrarrojo cercano (Vis-NIR; 400-2500 nm) detecta principalmente vibraciones de enlaces químicos de compuestos que contienen información sobre la composición inherente del suelo. Esta comprende minerales, compuestos orgánicos y agua (Stenberg et al., 2010; Viscarra Rossel & Behrens, 2010). Las absorciones amplias y débiles de longitudes de onda inferiores a 1000 nm pueden ser resultado de cromóforos (grupos funcionales) y óxido de hierro. Las absorbancias a longitudes de onda alrededor de 2200 nm ocurren a partir de minerales arcillosos y la MO absorbe en diferentes regiones a lo largo del rango Vis-NIR (Stenberg *et al.,* 2010). Otros rangos del espectro, más allá del Vis-NIR, generan más detalle, pero esto hace que la extracción de la información sea más desafiante (Viscarra Rossel *et al.,* 2006).

La espectroscopía Vis-NIR es una herramienta eficaz para el análisis de suelo, con ventajas como la predicción de varias propiedades del suelo a partir de una medición espectral (Metzger *et al.,* 2024), es rápida y fácil de usar, requiere poca

preparación de muestra, no es destructiva, es amigable con el medio ambiente y costo-efectiva (Nocita *et al.,* 2015). Por otro lado, la espectroscopía presenta desventajas como el control de las condiciones ambientales, ya que la humedad relativa alrededor del instrumento puede causar variaciones en los espectros (Rodríguez P *et al*., 2016), el uso de distintos equipos dificulta la comparaciones de datos y, además, se carece de protocolos globales para la preparación y lectura de las muestras (Poppiel *et al.,* 2022). Existen actualmente avances significativos en el desarrollo de la espectroscopía de suelos (Adeline *et al.,* 2017). Sin embargo, la falta de estandarización de los procesos limita la comparación de los resultados en las bases de datos globales. En ese sentido la Red Global de Laboratorios de Suelos (Global Soil Laboratory Network; GLOSOLAN; de su sigla en inglés) a través del programa GLOSOLAN Spectroscopy, pretende dar soporte para la estandarización de esta técnica. En esta iniciativa se analizan los procedimientos/protocolos de preparación de muestras, presentación de la misma al instrumento, estándares de control, entre otros, con el objetivo de aportar a la estandarización de los métodos de análisis (Poppiel *et al.,* 2022). Recientemente la FAO desarrolló un manual en el que se entrega información relevante para la preparación de las muestras y de su presentación al instrumento a utilizar (FAO, 2024).

El uso del Vis-NIR ha permitido estimar diversas variables del suelo. Entre las propiedades físicas evaluada se encuentran el contenido de arcilla (Islam et al.,

2003), limo y arena (Cozzolino & Morón, 2003). También se han estimado propiedades químicas como el contenido de carbono orgánico (Chang & Laird, 2002) y la materia orgánica (Daniel *et al.,* 2003), así como propiedades biológicas, incluyendo las actividades enzimáticas y la respiración de suelo (Soriano-Disla *et al.,* 2014). Además, Zornoza *et al.,* (2008) y otros autores sugirieron que la eficacia en el uso de C está fuertemente controlada por la comunidad microbiana. Las propiedades biológicas han sido las menos estudiadas a través de la espectroscopía, siendo un campo de exploración interesante para la comunidad científica. Es un desafío aún la estandarización de metodologías de análisis de las propiedades físicas del suelo, lo que dificulta la comparación de datos entre laboratorios.

Existe poca información relacionada con predicciones de propiedades biológicas en comparación con las propiedades físicas y químicas del suelo. Al desarrollar predicciones de propiedades biológicas, el pretratamiento de la muestra (por ejemplo: tamizado, preincubación y contenido de agua) y la variabilidad de datos de medición son consideraciones particularmente relevantes al no existir protocolos armonizados. El pretratamiento de las muestras es de gran relevancia en todas las medidas biológicas, la perturbación y las condiciones del pretratamiento pueden tener una fuerte influencia en el tamaño, composición y la actividad de los microorganismos en el suelo (Soriano-Disla *et al.,* 2014).

De los métodos de análisis de datos espectrales para espectroscopía de suelos, se destaca el ya clásico PLSR por su buen desempeño con la mayoría de las variables. Otros métodos como PCA-regresión, PCR-regresión han sido también parte de las opciones clásicas. Sin embargo, han ido surgiendo modernos métodos que están siendo incorporados por los analistas como Random Forest, Redes Neuronales entre otros.

El uso de la espectroscopia Vis-NIR requiere calibraciones mediante análisis multivariado, que permiten procesar datos y generar curvas de calibración de diversos atributos de los suelos en base a la información obtenida a partir de metodologías tradicionales. Una de las herramientas utilizadas para calibrar estos espectros Vis-NIR es la regresión de mínimos cuadrados parciales (del Inglés Partial Least Square Regression, PLSR) (Wold *et al.,* 1983). Este método de regresión PLS combina información de cientos de longitudes de onda en los primeros factores PLS (o variables latentes), mientras que los factores menos importantes pueden incluir efectos de fondo, no relevantes (Bolster *et al.,* 1996).

Hipótesis

Las propiedades bioquímicas del suelo en agroecosistemas pueden ser estimadas con alta precisión a través de espectroscopía Vis-NIR y modelos matemáticos.

Objetivos

Objetivo general

Evaluar la capacidad predictiva de modelos basados en análisis multivariados a partir de datos obtenidos por espectroscopía Vis-NIR para estimar propiedades bioquímicas del suelo en agroecosistemas del centro sur de Chile.

Objetivos específicos

- Analizar las propiedades bioquímicas del suelo (respiración, materia orgánica particulada (POM), biomasa microbiana y actividad β-glucosidasa) por métodos convencionales.

- Generar y validar modelos de predicción/estimación de variables bioquímicas a través de espectroscopía de suelos para el centro sur de Chile.

Referencias bibliográficas

Adeline, K. R. M., Gomez, C., Gorretta, N., & Roger, J.-M. (2017). Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma*, *288*, 143-153. https://doi.org/10.1016/j.geoderma.2016.11.010

Barra, I., Haefele, S. M., Sakrabani, R., & Kebede, F. (2021). Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. *TrAC Trends in Analytical Chemistry*, *135*, 116166. https://doi.org/10.1016/j.trac.2020.116166

Bolster, K. L., Martin, M. E., & Aber, J. D. (1996). Determination of carbon fraction and nitrogen concentration in tree foliage by near infrared reflectances: A comparison of statistical methods. *Canadian Journal of Forest Research*, *26*(4), 590-600. https://doi.org/10.1139/x26-068

Burbano-Orjuela, H. (2016). El suelo y su relación con los servicios ecosistémicos y la seguridad alimentaria. *Revista de Ciencias Agrícolas*, *33*(2), Article 2. https://doi.org/10.22267/rcia.163302.58

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J. T., & Reichstein, M. (2014). Global covariation of carbon turnover times with climate in terrestrial

ecosystems. *Nature*, *514*(7521), 213-217. https://doi.org/10.1038/nature13731

Chang, C.-W., & Laird, D. A. (2002). Near-infrared reflectance spectroscopic analysis of soil c and n. *Soil Science*, *167*(2), 110.

Clunes, J., Valle, S., Dörner, J., Martínez, O., Pinochet, D., Zúñiga, F., & Blum, W. E. H. (2022). Soil fragility: A concept to ensure a sustainable use of soils. *Ecological Indicators*, *139*, 108969. https://doi.org/10.1016/j.ecolind.2022.108969

Cotching, W. E. (2018). Organic matter in the agricultural soils of Tasmania, Australia – A review. *Geoderma*, *312*, 170-182. https://doi.org/10.1016/j.geoderma.2017.10.006

Cotrufo, M. F., & Lavallee, J. M. (2022). Chapter One - Soil organic matter formation, persistence, and functioning: A synthesis of current understanding to inform its conservation and regeneration. En D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 172, pp. 1-66). Academic Press. https://doi.org/10.1016/bs.agron.2021.11.002

Cozzolino, D., & Morón, A. (2003). The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *The Journal of Agricultural Science*, *140*(1), 65-71.

Daniel, K. W., Tripathi, N. K., & Honda, K. (2003). Artificial neural network analysis of laboratory and in situ spectra for the estimation of

macronutrients in soils of Lop Buri (Thailand). *Soil Research*, *41*(1), 47-59. https://doi.org/10.1071/sr02027

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Casanova Pinto, M., Casanova-Katny, A., Muñoz, C., Boudin, M., Zagal Venegas, E., & Boeckx, P. (2015). Soil carbon storage controlled by interactions between geochemistry and climate. *Nature Geoscience*, *8*(10), 780-783. https://doi.org/10.1038/ngeo2516

Islam, K., Singh, B., & Mcbratney, A. (2003). Simultaneous Estimation of Several Soil Properties by Ultra-Violet, Visible, and Near-Infrared Reflectance Spectroscopy. *Australian Journal of Soil Research - AUST J SOIL RES*, *41*. https://doi.org/10.1071/SR02137

Kazeev, K. Sh., Odabashian, M. Yu., Trushkov, A. V., & Kolesnikov, S. I. (2020). Assessment of the Influence of Pyrogenic Factors on the Biological Properties of Chernozems. *Eurasian Soil Science*, *53*(11), 1610-1619. https://doi.org/10.1134/S106422932011006X

Kawamura, K., Watanabe, N., Sakanoue, S., & Inoue, Y. (2008). Estimating forage biomass and quality in a mixed sown pasture based on partial least squares regression with waveband selection. *Grassland Science*, *54*(3), 131-145. https://doi.org/10.1111/j.1744-697X.2008.00116.x

Lavallee, J. M., Soong, J. L., & Cotrufo, M. F. (2020). Conceptualizing soil organic matter into particulate and mineral-associated forms to address

global change in the 21st century. *Global Change Biology*, *26*(1), 261-273. https://doi.org/10.1111/gcb.14859

Magdoff, F. (2001). Concept, Components, and Strategies of Soil Health in Agroecosystems. *Journal of Nematology*, *33*(4), 169-172.

Mann, C., Lynch, D., Fillmore, S., & Mills, A. (2019). Relationships between field management, soil health, and microbial community composition. *Applied Soil Ecology*, *144*, 12-21. https://doi.org/10.1016/j.apsoil.2019.06.012

McBratney, Alex. B., & Hartemink, A. E. (2024). Define soil. *Soil Security*, *14*, 100135. https://doi.org/10.1016/j.soisec.2024.100135

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J. A. M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., … Wetterlind, J. (2015). Chapter Four - Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. En D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 132, pp. 139-159). Academic Press. https://doi.org/10.1016/bs.agron.2015.02.002

NRCS, U. (2019). Soil Health Assessment | Natural Resources Conservation Service. Soil Health. United States Department of Agriculture Natural Resources Conservation

Poppiel, R. R., Paiva, A. F. da S., & Demattê, J. A. M. (2022). Bridging the gap between soil spectroscopy and traditional laboratory: Insights for routine

implementation. *Geoderma*, *425*, 116029. https://doi.org/10.1016/j.geoderma.2022.116029

Preethi, B., Poorniammal, R., Balachandar, D., Karthikeyan, S., Chendrayan, K., Bhattacharyya, P., & Adhya, T. K. (2013). Long-term organic nutrient managements foster the biological properties and carbon sequestering capability of a wetland rice soil. *Archives of Agronomy and Soil Science*, *59*(12), 1607-1624. https://doi.org/10.1080/03650340.2012.755260

Rodríguez P, W., García R, P. A., & Fajardo O, A. (2016). Aplicaciones de técnicas espectroscópicas para el análisis de suelos. *Revista Facultad de Ciencias Básicas*, *12*(2), Article 2. https://doi.org/10.18359/rfcb.2030

Sergeevna Kozun, Y., Shagidullovich Kazeev, K., & Ilich Kolesnikov, S. (2022). Climatic gradients of biological properties of zonal soils of natural lands. *Geoderma*, *425*, 116031. https://doi.org/10.1016/j.geoderma.2022.116031

Shaw, G. A., & Burke, H. K. (2003). *Spectral Imaging for Remote Sensing*. *14*(1).

Song, J., Shi, X., Wang, H., Lv, X., Zhang, W., Wang, J., Li, T., & Li, W. (2024). Improving Soil Quality Index Prediction by Fusion of Vis-NIR and pXRF spectral data. *Geoderma*, *447*, 116938. https://doi.org/10.1016/j.geoderma.2024.116938

Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., 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. https://doi.org/10.1080/05704928.2013.811081

Staff, S. S. (1999). *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys.* U.S. Department of Agriculture, Natural Resources Conservation Service.

Stark, C., Condron, L. M., Stewart, A., Di, H. J., & O'Callaghan, M. (2007). Influence of organic and mineral amendments on microbial soil properties and processes. *Applied Soil Ecology*, *35*(1), 79-93. https://doi.org/10.1016/j.apsoil.2006.05.001

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Chapter Five—Visible and Near Infrared Spectroscopy in Soil Science. En D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 107, pp. 163-215). Academic Press. https://doi.org/10.1016/S0065-2113(10)07005-7

Telo da Gama, J. (2023). The Role of Soils in Sustainability, Climate Change, and Ecosystem Services: Challenges and Opportunities. *Ecologies*, *4*(3), Article 3. https://doi.org/10.3390/ecologies4030036

Valle, S. R., & Carrasco, J. (2018). Soil quality indicator selection in Chilean volcanic soils formed under temperate and humid conditions. *CATENA*, *162*, 386-395. https://doi.org/10.1016/j.catena.2017.10.024

Viscarra Rossel, R. A., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, *158*(1), 46-54. https://doi.org/10.1016/j.geoderma.2009.12.025

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, *131*(1), 59-75. https://doi.org/10.1016/j.geoderma.2005.03.007

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, *131*(1), 59-75. https://doi.org/10.1016/j.geoderma.2005.03.007

Walsh, E., & McDonnell, K. P. (2012). The influence of added organic matter on soil physical, chemical, and biological properties: A small-scale and short-time experiment using straw. *Archives of Agronomy and Soil Science*, *58*(sup1), S201-S205. https://doi.org/10.1080/03650340.2012.697999

Wold, S., Martens, H., & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. En B. Kågström & A. Ruhe (Eds.), *Matrix Pencils* (pp. 286-293). Springer. https://doi.org/10.1007/BFb0062108

Zabaloy, M. C. (2021). Una sola salud: La salud del suelo y su vínculo con la salud humana. *Revista Argentina de Microbiología*, *53*(4), 275-276. https://doi.org/10.1016/j.ram.2021.11.001

Zhang, J., Li, Y., Jia, J., Liao, W., Amsili, J. P., Schneider, R. L., van Es, H. M., Li, Y., & Zhang, J. (2023). Applicability of soil health assessment for wheat-maize cropping systems in smallholders' farmlands. *Agriculture, Ecosystems & Environment*, *353*, 108558. https://doi.org/10.1016/j.agee.2023.108558

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K. M., Arcenegui, V., & Mataix-Beneyto, J. (2008a). Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biology and Biochemistry*, *40*(7), 1923-1930. https://doi.org/10.1016/j.soilbio.2008.04.003

Capítulo 1 To be submitted to Geoderma Regional

The potential of Vis-NIR spectroscopy to estimate biochemical soil properties in agroecosystems of the South-Central zone of Chile

Jean Intriago-Ávila[1], Erick Zagal[2*], Lizardo Reyna[3], María de los Ángeles Sepulveda-Parada[2], Miguel Garriga[4], María Luisa Izaguirre-Mayoral[3], Camila Ramos[5]

[1] Master Program in Agronomic Sciences, Faculty of Agronomy, Universidad de Concepción, Chillán 3812120, Chile; jintriago2023@udec.cl

[2] Spectroscopy Laboratory (Vis-IR) and Sustainable Soil Management, Soils and Natural Resources Department, Faculty of Agronomy, University of Concepción, Chillán 3812120, Chile; ezagal@udec.cl; mariadelosasepu@udec.cl

[3] Faculty of Agricultural Engineering, Technical University of Manabí, 13132, Ecuador; lizardo.reyna@utm.edu.ec; maria.izaguirre@utm.edu.ec

[4] Department of Plant Production, Faculty of Agronomy, University of Concepción, Chillán 3812120, Chile; mgarriga@udec.cl

[5] Doctoral Program in Agronomy Sciences, Faculty of Agronomy, Universidad de Concepción, Chillán 3812120, Chile; camiramos@udec.cl

* Corresponding author: ezagal@udec.cl

Abstract

The use of visible near-infrared spectroscopy (Vis-NIR) presents a non-destructive, fast, reliable, and cost-effective approach to predict soil physical, chemical and biochemical properties. In this study, biochemical properties such as basal soil respiration, active microbial biomass (Fluorescein diacetate hydrolysis), β-glucosidase activity and particulate organic matter-C (POM), key variables in the carbon cycle, were estimated. Seventy sites were sampled in south-central Chile, specifically in the regions of Maule, Biobío, Ñuble and Araucanía (35°58′ S 72°58′ W) in traditional crops, pastures, and fruit trees at a depth of 0-30 cm. The samples were analyzed in a laboratory and scanned using a UV-Vis-NIR spectrophotometer covering a range from 175 to 3300 nm and a resolution of 1 nm. Multivariate models including PLS, a linear regression method, and Random Forest (RF), which is based on machine learning, were used. Seven preprocessing methods were evaluated, and the orthogonal signal correction (OSC) emerged as the best preprocessing technique in PLS modelling. PLS models, with preprocessing OSC, and multiplicative scatter correction (MSC), showed high accuracy on traits such as basal respiration and β-glucosidase activity (R^2validation: 0.99-0.98; RMSECV: 0.05-0.101, respectively). The FDA required logarithmic transformations, while POM was sensitive to spectral corrections (R^2validation: 0.95; RMSECV: 1.0486). Random Forest (RF) presented heterogeneous results: POM was well predicted (RPD = 2.08; R^2validation = 0.77, RMSEP:1.128), but biomass and β-glucosidase activity had

low performances. Based on these results, soil spectroscopy has great potential as a tool for the estimation of soil biochemical attributes.

Keywords: latitudinal gradient, soil spectroscopy, PLS, orthogonal signal correction, random forest, particulate organic matter-C.

Introduction

Chile boasts a remarkable diversity of soils, encompassing 11 of the 12 soil orders recognized by the United States Department of Agriculture (Staff, 1999). This diversity is intricately linked to the country's varied natural agroecosystems, which have spurred the development of numerous human-managed production systems. Soil is widely regarded as one of the most vital natural resources, given its critical roles in food production, water regulation, organic carbon and nutrient cycling, and providing habitat for biodiversity. Indeed, the quality or health of soil is fundamental to sustaining these essential ecosystem functions and services (Montanarella *et al.,* 2015).

Soil organic carbon (SOC) represents one of the largest terrestrial carbon reservoirs. Its atmospheric exchange is influenced by climatic conditions and the soil's geochemical properties, both of which play a crucial role in stabilizing carbon within the mineral matrix (Doetterl *et al.,* 2015). SOC plays a critical role in agroecosystems by supporting microbial activities and serving as a key indicator of soil health, owing to its significant contributions to soil quality and climate

change mitigation. Research has shown that soils with higher biological activity typically exhibit greater microbial biomass, leading to increased soil enzyme activity and elevated basal respiration rates (Stark *et al.,* 2007). Soil enzymes enable microorganisms to access energy and nutrients by breaking down complex substances through degradation and mineralization, as well as by forming complex, stable molecules during humification. Often, soil enzyme activities correlate with native SOC levels, reflecting larger microbial communities and the stabilization of enzymes in humic materials. In particular, β-glucosidase activity is commonly used to gauge changes in both the quantity and quality of SOC (Preethi *et al.,* 2013).

Organic matter (OM) is an organic fraction that influences the physical, chemical and biological properties of the soil, and constitutes a reservoir of nutrients, which by the action of soil microorganisms is actively decomposed (Walsh & McDonnell, 2012). Soil organic matter is a dynamic and changing resource that reflects the balance between the addition of new OM and the loss of OM already in the soil, which is partly controlled by living biological activity (Cotching, 2018). Soil OM is fundamental to many ecosystem services, and a recent conceptual model divides it into two distinct reservoirs with different properties and behaviors: particulate organic matter (POM) and matrix-associated organic matter (MAOM) (Cotrufo & Lavallee, 2022). POM primarily consists of light fragments of relatively undecomposed plant material, which exhibit higher activation energy. POM is

more sensitive to climate change compared to MAOM, which benefits from protection through its mineral associations in the soil (McClure, 2003; Skjemstad *et al.,* 2004; Poeplau & Don, 2013). Due to this difference in susceptibility to decomposition, MAOM generally persists in soil much longer than POM (Lavallee *et al.,* 2020).

Obtaining accurate and rapid information about the soil is essential for its sustainable and efficient use. Soil is a heterogeneous system in which the underlying processes are complex and not totally understood (Viscarra Rossel *et al.,* 2006). Traditionally, gathering soil data has involved labor-intensive and costly procedures, including field sample collection followed by extensive laboratory preparation and analysis (e.g., wet chemistry), which relies on chemical reagents and specialized equipment (Wang *et al.,* 2023). One of the major challenges imposed by climate change is the need to promptly and accurately assess soil properties. Continuous monitoring of the soil's chemical, physical, and biological attributes is crucial to determining its health for agricultural productivity and for supporting other ecosystem services. To generate information more quickly and efficiently, several innovative methodologies have been developed, with soil spectroscopy emerging as one of the most promising techniques over the past two decades.

This powerful analytical method relies on the interaction between matter and electromagnetic radiation at specific frequencies, resulting in a unique spectral signature (Shaw & Burke, 2003). Infrared spectroscopy particularly in the visible and near-infrared (Vis-NIR, 400–2600 nm) ranges, has emerged as a fast, reliable technique for obtaining detailed soil information (Padarian *et al.,* 2019; J. Yang *et al.,* 2020; Zhong *et al.,* 2021). Its non-destructive nature, high throughput that enables rapid evaluation of numerous soil samples (Viscarra Rossel *et al.,* 2022), minimal toxic waste generation, ease of sample storage, and overall cost-effectiveness make it an attractive option for modern soil analysis.

When mid-infrared (MIR) or NIR energies are applied to the soil, light scatters within the sample, causing molecular bonds to vibrate and absorb specific wavelengths. The remaining light is diffusely reflected back to a detector, which records the response as a function of wavelength (Griffiths, 2010). These vibrations promote the molecular bonds from their lowest energy state to their first excited state. In contrast, harmonics and combination bands arise when transitions occur from the fundamental state to the second or higher vibrational states or when energy is exchanged between two or more vibrational modes; NIR spectra primarily result from these harmonics and combined vibrations. In the visible range, the higher energy of the light leads to electron excitation and electronic transitions (Picollo *et al.,* 2019).

Depending on the sensor's resolution, a Vis-NIR or MIR soil spectrum can contain hundreds to thousands of frequencies, providing valuable information about various soil properties. For example, soil color and iron oxides, such as hematite and goethite, exhibit characteristic absorption features at 400, 450, 500, 650, and 900 nm. Clay minerals, including gibbsite and kaolinite, display distinct absorption bands at 1400 and 2200 nm, while illite and smectite show bands at 2200, 2340, and 2445 nm, and 2,200 nm, respectively. Carbonates are typically detected at 2335 nm, and gypsum can also be identified within this range. Organic matter types exhibit absorption features at 1100, ,600, 1700, 1800, 2000, and between 2200 and 2400 nm. Additionally, water content, both adsorbed and free, is detectable at 1,400 and 1,900 nm, while particle size can also influence spectral characteristics (Nguyen *et al.,* 1991; Viscarra Rossel & Hicks, 2015).

Several studies have demonstrated the effectiveness of Vis-NIR spectroscopy in predicting soil biological properties. For instance, Zornoza *et al.,* (2008) successfully applied this technique to estimate enzymatic activities (e.g., β-glucosidase) and basal soil respiration, with high accuracy. Similarly, Dai *et al.,* (2025) reported favorable correlations when predicting POM fractions using Vis-NIR spectroscopy. Additionally, specific wavelengths have shown potential in enhancing these estimations. Vohland *et al.,* (2014) have documented successful predictions of microbial biomass carbon and nitrogen using this approach. Despite

being Vis-Nir spectroscopy a promising alternative there are some biological parameters such as FDA where the information is scarce.

Herrmann *et al.,* (2009) implemented Vis-NIR spectroscopy combined with a partial least squares regression (PLSR) to predict active microbial biomass in compost. Their cross-validated model achieved an R^2 = of 0.53, highlighting its potential for practical applications in soil biological assessments. On the other hand, Vis-NIR allows digital soil mapping applications.

Recent research by X. Li *et al.,* (2024) highlighted that MIR range is gaining prominence in biochemical property predictions. Table 1 shows several biological traits estimated using Vis-NIR and MIR spectroscopy.

An essential step in data manipulation before modeling is data preprocessing (Rinnan *et al.,* 2009). These techniques modify the spectral signal by eliminating irrelevant information and enhancing model robustness (Gobrecht *et al.,* 2014). Spectral pre-processing methods are also applied to linearize the often nonlinear relationship between spectral data and soil properties (Xu *et al.,* 2021). Additionally, they improve absorption characteristics and minimize physical effects and noise in spectroscopic measurements. Among the various spectral preprocessing methods are continuum elimination, multiplicative scatter correction (Clark & Roush, 1984), first derivative transformation, and standard

normal variate normalization (Barnes *et al.,* 1989). Orthogonal Signal Correction (OSC) has limited application in soil science (Wold *et al.,* 1998). However, actual literature using OSC has demonstrated that this pre-treatment improved accuracy of soil organic carbon in PLSR (Biney *et al.,* 2021). The primary objective of preprocessing is to remove variations unrelated to the predicted parameter while preserving essential information that might be lost using other methods.

In the case of OSC, it is assumed that any variation not linked to the response variable is an artifact that must be filtered out without compromising critical information. This ensures the output component remains mathematically orthogonal to the response or as close to orthogonality as possible (Engel *et al.,* 2013; Wold *et al.,* 1998).

Developing an accurate model is essential to minimize errors. To date, partial least squares regression (PLSR) has been the most widely used method for processing Vis-NIR spectra (Greenberg *et al.,* 2022; S. Li *et al.,* 2015). The significant advancements in computer capacity and processing speed over the past two decades have allowed mathematicians and data scientists to develop more sophisticated Machine Learning (ML) algorithms.

Recently, the development of ML algorithms has enhanced the modeling of soil properties using spectral data (Mouazen *et al.,* 2010; Rossel & Behrens, 2010).

The ML, a subfield of artificial intelligence, includes various algorithms for predictive modeling through pattern recognition, such as Random Forest (RF) (Breiman, 2001). Several ML algorithms, including RF, have been successfully applied in soil spectroscopy (Tan *et al.,* 2020; Vašát *et al.,* 2017). ML algorithms offer several advantages over linear models, such as the ability to model nonlinear relationships between variables and the absence of normality assumptions for predictor variables (Brown *et al.,* 2000). These characteristics can lead to improved accuracy in ML models, especially in complex feature spaces, when compared to linear algorithms (Rodriguez-Galiano *et al.,* 2012).

This study hypothesizes that soil biochemical properties in agroecosystems of Central-Southern Chile can be accurately estimated using Vis-NIR spectroscopy combined with mathematical models. The primary objective is to assess the predictive capacity of multivariate analysis-based models and Machine Learning (ML) algorithms for soil biochemical properties, as determined by conventional analytical methods.

Materials and methods

Soil Sampling

Soil sampling was conducted during two periods-between March and June 2022 and between April and July 2024 in the Maule (35°58′ S, 72°17′ W), Ñuble (36°00″

S, 71°57″ W), Biobío (36°48″ S, 71°38″ W), and Araucanía (37°46″ S, 72°58″ W) regions of Chile, encompassing a total of 70 sites (Figure 1).

Site descriptions

The climatic characteristics of Chile are strongly influenced by its coastal and Andean Mountain ranges. In the central-southern region where sampling was conducted, the annual average temperature ranges from 0°C to 15°C, with summer temperatures reaching up to 40°C. Annual rainfall varies between 350 mm and 1000 mm. Most of the soils along the transect are derived from volcanic ash due to the intense volcanic activity during the Quaternary period and are rich in organic carbon (Valle & Carrasco, 2018). These edaphoclimatic conditions are typical of the Mediterranean zone.

The transect was selected to maximize the climatic, physical, chemical, and biological diversity of soils, thereby enabling the development of models with broad applicability (Sepúlveda *et al.,* 2021). The study area includes soils originating from various parent materials (e.g., volcanic ash, alluvial deposits, marine sediments) and encompasses diverse geological environments (e.g., continental sedimentary, plutonic, metamorphic). It also covers a range of geomorphological settings from fluvial or alluvial sedimentation plains and central fluvio-glacio-volcanic plains to marine or fluvial-marine plains as well as varied

physiographic positions (mid-slope, plain, plateau, high slope, valley bottom, low slope, terrace) and topographies (flat, undulating, and rugged) (Table S1).

Samples were collected at depths of 0–30 cm using a Dutch auger. An “X” sampling method was employed by drawing imaginary cross-shaped lines across the field and taking five subsamples along these lines, which were then homogenized to form a composite sample. The samples were stored at 4°C in a controlled cooler during transport to the University of Concepción, where they were analyzed at the VIS-IR Spectroscopy and Sustainable Soil Management Laboratory of the Faculty of Agronomy. The physical and chemical properties of soil samples are summarized in Table 2.

Soil biological properties assessment

Prior to analysis, samples were sieved through a 2 mm mesh to remove gravel and roots. However, some samples with high moisture content could not be sieved immediately and were first air dried at room temperature. Any portions of soil that dried during this process were discarded.

According to Barajas *et al.,* (1994), each stage of sample preparation must be performed rigorously to ensure accurate and comparable results, while also preventing metabolic alterations in microorganisms. For soil sample conditioning,

the bulk density was determined using the cylinder method (Ellies, 1990)**,** the real density was measured with the pycnometer method as described by Blake & Hartge (1986), and a subsample was taken to determine gravimetric moisture (ISO 11465, 1993). In addition, the soil samples were conditioned to evaluate the water-filled pore space (WFPS), a metric defining the proportion of pore space occupied by water, which directly influences biological activity. Typically, properties related to aerobic processes are assessed at 60% WFPS, with samples incubated for 7 days at 20°C to reactivate microbial activity (Paul, 2007).

After adjusting soil moisture to 60% WFPS, 15 grams of soil were placed in 50 mL Falcon tubes (in triplicate), then hermetically sealed in triplicate and incubated at 22°C for 7 days. Measurements of respired CO_2 were recorded on days 1, 4, and 7 using a Li-820 CO_2 analyzer (Li-COR Bioscience, Lincoln, NE, USA). The analyzer was calibrated with a standard CO_2 gas at 1073.75 ppm, delivered as a laminar nitrogen flow (UHP; 99.99%; Linde; Chile). CO_2 measurements were obtained by sampling the headspace with a 1 mL syringe, and LI-COR software enabled real-time monitoring of peak heights and cell output at 1-second intervals (Sherrod *et al.,* 2012). This process measures the oxidation of organic matter by aerobic microorganisms, which produces carbon dioxide (CO_2) as an end product (Alef, 1995).

FDA was performed following Alef, (1995). Briefly, soil samples were placed in screw-capped test tubes and mixed with a 60 mM sodium phosphate buffer, fluorescein diacetate, and hydrolyzed fluorescein. After vigorous vortexing, the samples were incubated at 25°C for 1 hour. Samples were then cooled in an ice bath, and acetone was added. The supernatant was transferred to Eppendorf tubes (1.5 ml) and centrifuged at 6000 rpm for 5 minutes at room temperature using a Lab-Tec Mikro 220R centrifuge (Föhrenstr, Tuttlingen, Germany). The absorbance of the supernatant was measured at 490 nm to quantify the overall active microbial activity in the soils (Schnürer & Rosswall, 1982).

Since β-glucosidase plays a crucial role in the degradation of soil carbohydrates—its hydrolytic products serving as a significant energy source for soil microorganisms—its activity is often correlated with the soil's organic carbon content (Eivazi & Tabatabai, 1988; Tabatabai, 1982). The β-glucosidase activity was quantified using the p-nitrophenol method. Soil samples were incubated with the β-D-glucopyranoside at a concentration of 25 mM in a buffered medium (MUB-HCl, pH 6) at 37°C for 1 hour. Following incubation, the samples were cooled in an ice bath, and the released p-nitrophenol was extracted by adding 0.5 M $CaCl_2$ and THAM-NaOH buffer (pH 12), then centrifuged at 6000 rpm for 5 minutes at room temperature. The absorbance of the supernatant was measured at 400 nm. Both the FDA and β-glucosidase assays were performed using a UV-1601 spectrophotometer (Beijing Reyleigh Analytical Instrument Co., Ltd., China).

Organic matter fractionation

The fractionation of soil organic matter was performed following the method described by Lavallee *et al.,* (2020). Briefly, 5 g of oven-dried soil was dispersed in a 0.5% sodium hexametaphosphate solution and shaken for 18 hours with glass beads to ensure complete dispersion. The resulting suspension was then passed through a 53 µm sieve. The fraction that passed through the sieve (<53 µm) was collected as mineral-associated organic matter (MAOM), while the material retained on the sieve was designated as particulate organic matter (POM). After drying at 60°C to a constant weight, each fraction was recovered and weighed to determine the percentage of recovery. The dried soil samples were then finely ground using a mortar until a fine powder consistency was achieved. Subsequently, carbon and nitrogen concentrations were analyzed using an elemental analyzer (LECO Corporation, 300 Lakeview Ave, USA), to assess carbon stocks in the top-soil layer.

Soil spectral measurements

Prior to spectral scanning, 200 grams of soil were sieved through a 2 mm mesh. The samples were then air-dried at room temperature before being oven-dried at 35°C for 24 hours. After drying, they were homogenized for 8 hours using an analog orbital shaker (Reax 20, HEIDOLPH®, Instruments, GmbH & Co. K, Schwabach, Germany).

Spectral scanning was conducted in the UV-Vis-NIR range (175–3300 nm) at 1 nm intervals using an Agilent Cary 5000 spectrophotometer (Agilent Technologies, USA). Samples were placed in a 200 µL sample holder, and each underwent 10 repetitions, including extraction, homogenization, and subsequent measurement in spectrophotometer. Spectral reflectance was measured at room temperature under dark conditions provided by the spectrophotometer to minimize ambient light interference. Before measurements, the instrument was calibrated using a white reference panel (Agilent Technologies, USA).

Spectral preprocessing

The spectral curves were subjected to various types of preprocessing to remove noise in the signal and highlight spectral features due to sample properties. The Savitzky-Golay algorithm with a window size of 11 and polynomial order 1 was initially used to smooth the curve and accentuate the shape of each spectral signature (Rinnan et al., 2009). This process was performed using Python programming (Python Software Foundation, https://www.python.org/), and its libraries for data analysis. The extremes of the curves were discarded due to signal leaving the curves withing the range of 190 to 2690 nm.

Data were preprocessed by mean-centered, multiplicative scatter correction (MSC) (Clark & Roush, 1984), second derivative, normalized (Barnes *et al.,* 1989), baseline correction (Buddenbaum & Steffens, 2012), transformation to log10, smooth removing outliers using the Pirouette® software (1030 N State ST APT 40C, Chicago, Illinois, United States). Then the OSC was applied along with additional preprocessing considering the cumulative variance and the number of factors or latent variables (LV).

Development of predictive models

Partial least squares regression (PLSR)

Partial least squares regression (PLSR) is a widely used linear regression method that combines principal component analysis (PCA) with multiple linear regression for statistical analyzing of multivariate data. A key aspect of is effective application is to selecting the optimal number of factors for the regression; using too many factors can lead to overfitting, while too few can result in underfitting (T. Zhang *et al.,* 2024). The primary objective of PLSR is to extract the maximum explanatory power from the dependent variable (Y) reducing dimensionality and eliminating noise (Wold *et al.*, 2001). The validation of the developed models was carried out using the 'leave one out' cross validation method as an internal validation. This procedure consists of extracting a sample from the training set and performing a partial least squares (PLS) regression on the remaining

samples. Then, the value of the omitted sample is predicted, and the corresponding error is analyzed. This process is repeated until each sample has been left out once. The model performance was determined based on the coefficient of determination (R^2) and the root mean square error of cross-validation (RMSECV) (Shao, 1993).

Random Forest (RF)

Random Forest (RF) is an ensemble machine learning method used for classification and regression, known for its model stability through tree diversity. It offers fast training, ease of implementation, and the ability to assess feature importance. RF constructs multiple decision trees using bootstrapping, where samples are randomly selected with replacement from the calibration set During tree construction, a subset of features is randomly chosen at each split node (typically between 1 and 10). The number of trees (ntree), commonly set to 500 for stability, and the number of selected features are optimized based on the lowest root mean square error (RMSE) from five-fold cross-validation, performed exclusively on the calibration set with randomly selected folds (Liaw & Wiener, 2002). Samples not used during training, known as "out-of-bag" (OOB) samples, serve to evaluate individual tree performance. By averaging the error across all trees, RF provides an unbiased internal estimate of generalization error (RMSEP) (Rodriguez-Galiano *et al.,* 2014). A low generalization error means the model has

a good ability to accurately generalize and predict new data. A high generalization error means the model overfits the training data and cannot accurately predict new data.

RF- tree numbers

The development of the RF algorithm was performed with the scikit-learn Python library in the JypiterLab environment (https://jupyter.org/). The first step was to load the spectral curves and perform a principal component analysis to identify clusters and detect extreme data. Then, cross-validation was performed (k-fold=5) using an increment of tree numbers from 100 to 500. This procedure allows us to select the appropriate number of trees for the model. Quality statistic parameters of the model performance, including, the RMSE, RDP (ratio of standard deviation) and the coefficient of determination (R^2) were used (**¡Error! No se encuentra el origen de la referencia.**). Cross-validation generates two graphs representing the quality indicators as a function of the number of trees, where the user can visually determine the appropriate number of trees for the model. The cross validation generates two graphs representing the quality indicators as a function of the number of trees where the user can visually select the appropriate number of trees for the model.

Subsequently, the data sets were divided into two groups, 80% for calibration and 20% for validation, and an iterative procedure was performed to identify the optimal split (80/20%) in terms of predictive performance (Reyna *et al.,* 2017). A reproducible pseudo-random split of the data was performed with each iteration, in selecting the best calibration/validation dataset based on the coefficient of determination (R^2) obtained using RF models, following the methodologies established in previous studies (Mallah *et al.,* 2022).

For the final models based on RF, the quality of the predictions was evaluated using the RMSEP and R^2, which quantifies the average magnitude of the prediction errors, these indicators allow a robust and comparative evaluation of the predictive performance. It is important to note that the calibration and validation set share the same data distribution. The prediction of the final models was evaluated using the RMSEP, RDP and R^2.

Model evaluation

For model evaluations four equations were used: Equation 1, 2, 3 and 4. See at the end of Tables and Figures

Results

Biochemical parameters

Biochemical analysis results for soil respiration, β-glucosidase activity, microbial biomass (FDA) and particulate matter-C are shown in Table 3. A moderate to high variability was observed in all traits (e.g., POM-C with SD = 2.35% and range = 0.05-10.53 %). To build robust prediction models, it is key that calibration values are not concentrated at extremes (min/max) but are close to the mean/median (e.g. soil respiration: mean = 0.85, median = 0.83). This reduces the extrapolation error, as spectral models predict better within the range with higher data density (Figure 2).

Spectroscopic analysis result

The PLS model, together with the preprocessing techniques, showed high predictive ability in biochemical variables after cross-validation whit cumulative variance ranging from 79.1% to 99.6% and prediction errors from 0.05 to 10.205 (Table 4). Models for Basal soil respiration and β-glucosidase activity achieved high prediction power with preprocessing including OSC, MSC and normalization with R^2 val of 0.99 and 0.98, and RMSECV: 0.05-0.101, respectively

The model for FDA required logarithmic transformations with baseline correction to maximize its performance (R^2 val = 0.92), while the development of models for

POM-C used spectral methods such as OSC and baseline correction. It was noted that preprocessing, such as second derivative or smoothing, reduced predictive performance generation in some cases (e.g., soil respiration with cross-validation the R^2 dropped from 0.99 to 0.55), suggesting risks of overfitting by introducing unneeded complexity.

The RF model was evaluated by external validation and presented a heterogeneous performance (Table 5). Only the POM-C showed a robust model, with an RPD of 2.08 (R^2 val = 0.77; RMSEP = 1.128), and good predictions were found with RPD>2. In contrast, the predictive capacity for microbial biomass: RPD = 1.30 and β-glucosidase: RPD = 1.34) is limited or poor with an RPD < 1.5, with moderate to-low power prediction (R^2 val= 0.35-0.45). The PLS model is cross-validation metrics, indicating a strong internal fit; however, it does not guarantee external generation. The RF model was validated using independent samples and exhibited lower overall performance except in POM-C where both approaches were comparable (PLS: R^2 val = 0.95 and. RF: R^2 val = 0.77).

Discussion

The choice of model depends on user considerations, including the type of validation (internal or external) and the nature of the data. For example, PLS is

typically suited for linear systems, while RF is preferred for variables with complex interactions.

The Vis-NIR spectroscopy technique proved to be a versatile tool for predicting soil biochemical variables. In this study, the results for soil respiration, β-glucosidase activity and POM-C revealed a recurring pattern: PLS models with targeted preprocessing techniques (e.g., OSC) achieved exceptional accuracy in cross-validation (R^2 = 0.95-0.99). For example, basal soil respiration showed an R^2 val = 0.99 with OSC + MSC + Normalization, exceeding even the R^2 val= of 0.86 reported by Gandariasbeitia *et al.,* (2017) for wet samples. Biney *et al.,* (2021) have also shown the valuable use of OSC in predicting, with excellent performance, SOC in a temporal gradient. Cross-validation is a calibration training technique that is particularly valuable when data are limited, as in this study. It is commonly used to validate and fine-tune models during training by assessing performance and determining the optimal number of factors, without needing an independent validation set. In practice, the method involves removing one or more samples from the calibration set and predicting their values using models built from the remaining data. Although cross-validation tends to yield an optimistic estimate of model performance, especially with small sample sizes it may not fully capture how the model will perform on truly independent data (Dardenne *et al.,* 2000). Therefore, it is advisable to evaluate the model using an independent test suite for a more accurate performance assessment. Indeed, due to the small

number of sites, external validation results in unacceptable prediction values (data not shown). In the case of the RF model, an external validation was performed, which yielded values ranging from R^2 = 0.35 to 0.77, demonstrating low to acceptable performance. For both models, this reflects the intrinsic challenges of predicting soil properties linked to transient biological dynamics and spatial heterogeneity. For β-glucosidase activity, a similar behavior was observed. When PLSR with OSC+MSC+Normalized was used, an R^2 val = 0.98 was obtained, close to the 0.96 reported by Cohen *et al.,* (2005), highlighting that spectral signals associated with microbial processes are highly sensitive to changes in humidity. The explanation lies in the dual nature of this enzyme: while its association with organic matter generates stable spectral signatures in NIR, its activity depends on variables such as pH or the presence of inhibitors, which lack a direct spectral fingerprint and limit generalization. Similarly, a good performance for POM-C was found in this study with PLSR and the pretreatment OSC (R^2 val = 0.95). Other researchers have reported R^2 values for cross validation somewhat lower, e.g Yang *et al.,* (2012) found an R^2 val = 0.92 and Ramifehiarivo et al., (2023) a R^2 val of 0.74. This is attributed to the PLS model, which models global correlations between soil organic matter (SOM) between bands (e.g., 1700-2200 nm). However, Das *et al.,* (2023) warn about the possibility of a drop when performing external validation where their model was very poor (R^2 of 0.10 for RF in labile carbon), but the results of the present study carried out by us in RF demonstrated better performance.

Recommendations by Terhoeven-Urselmans *et al.,* (2008) mention that moisture in wet samples masks informative spectral bands (e.g.1450-1950 nm) or introduces physical variability (e.g., moisture gradients within the samples) while drying, reducing parameters such as ergosterol a crucial component of fungal cell membranes, homogenizes reflectance. This duality requires standardized protocols that balance biological preservation and spectral consistency, a methodological gap still unresolved. In addition, Sankey *et al.,* (2008) reported that local-scale models based on site-specific data help to obtain better results in soil predictions but require a larger or higher number of samples. However, global or regional models, although economically efficient, have limitations in accuracy. This indicates that the variability of soil and soil types used in the model directly influences model performance.

Recent research by Li *et al.,* (2024) highlighted that, compared to Vis-NIR spectra, the peaks and valleys in MIR spectra are primarily attributed to energy level transitions caused by fundamental vibrations. As a result, MIR spectra provide richer information than Vis-NIR spectra. Currently, the MIR range is gaining prominence in biochemical property predictions, while Vis-NIR remains the leading choice for digital soil mapping applications. This demonstrates the need to expand the spectral range for the estimation of biological properties. The

literature shows a few studies where the use of both techniques for this type of properties can be compared (Table 1).

Conclusion and Perspective

Vis-NIR spectroscopy proved to be a promising tool for predicting soil biochemical variables (Basal soil respiration, β-glucosidase, Active microbial biomass and POM-C) under controlled conditions reaching techniques R^2 predictions 0.90 using multivariate models and data pre-processing techniques such as OSC. However, its applicability in real scenarios faces limitations because of soil moisture which is fundamental in soil biological processes. The RF algorithm showed acceptable performance for some parameters (POM-C) with external validation, demonstrating its validity and utility as a suitable nonlinear algorithm for this type of prediction. The gap between internal fit and generalization highlights the need for rigorous validation and standardized protocols. While cross-validation for PLS has shown very good results, due to the inherent variability of the geoclimatic gradient used and for the extensive use of the generated models, acceptable external validation is required. This can be achieved by i) expanding the modeling dataset (spectral libraries or sampling), ii) exploring the clustering of soils with similar formation factors (e.g., volcanic soils), and iii) generating local or regional models. On the other hand, the application of

OSC in soil science is a very promising preprocessing that should be included in soil spectroscopy-related studies.

The global scientific community, led by Peng *et al.,* (2025), is moving towards the implementation of soil spectroscopy to obtain digitized soil data. This effort, comprised of the International Union of Soil Sciences (IUSS) and GLOSOLAN-Spec, aimed to address seven key challenges hindering the global adoption of soil spectroscopy.

Declaration of Conflict of Interest

There are no conflicts of interest

Acknowledgments

This study was funded by the Fondef Project ID21I10264 entitled "Development and Validation in soil analysis of a cost-effective, rapid and environmentally friendly process at local and global level: the alternative of soil spectroscopy." Authors would like to thank the Spectroscopy Laboratory (VIS-IR) and Sustainable Soil Management, and Ms. Katherine Rebolledo for their valuable technical support. Thanks to the master's program in Agricultural Sciences.

Declaration of authorship contribution

Jean Intriago-Ávila: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Erick Zagal:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing**. Lizardo Reyna:** Methodology, Software, Validation, Visualization, Writing – review & editing. **María de los Ángeles Sepúlveda-Parada:** Methodology, Software, Validation, Visualization, Writing – review & editing. **Miguel Garriga:** Visualization, Writing – review & editing. **María Luisa Izaguirre-Mayoral:** Validation, Visualization. **Camila Ramos:** Writing – review & editing.

References

Adeline, K. R. M., Gomez, C., Gorretta, N., & Roger, J.-M. (2017). Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma*, *288*, 143-153. https://doi.org/10.1016/j.geoderma.2016.11.010

Alef. (1995). *Methods in applied soil microbiology and biochemistry* (Alef y Nannipieri (1995) (Eds.)). Academic Press, London.

Barajas, M., Grace, C., Hart, M., Lin, Q., & Brookes, P. (1994). *Laboratory Manual of the Soil Microbial Biomass Group, Soil Science Dept.* https://doi.org/10.13140/RG.2.1.3911.2407

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, *43*(5), 772-777. https://doi.org/10.1366/0003702894202201

Barra, I., Haefele, S. M., Sakrabani, R., & Kebede, F. (2021). Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. *TrAC Trends in Analytical Chemistry*, *135*, 116166. https://doi.org/10.1016/j.trac.2020.116166

Biney, J. K. M., Blöcher, J. R., Borůvka, L., & Vašát, R. (2021). Does the limited use of orthogonal signal correction pre-treatment approach to improve the prediction accuracy of soil organic carbon need attention? *Geoderma*, *388*, 114945. https://doi.org/10.1016/j.geoderma.2021.114945

Blake, G. R., & Hartge, K. H. (1986). Particle Density. En *Methods of Soil Analysis* (pp. 377-382). John Wiley & Sons, Ltd. https://doi.org/10.2136/sssabookser5.1.2ed.c14

Bolster, K. L., Martin, M. E., & Aber, J. D. (1996). Determination of carbon fraction and nitrogen concentration in tree foliage by near infrared reflectances: A comparison of statistical methods. *Canadian Journal of Forest Research*, *26*(4), 590-600. https://doi.org/10.1139/x26-068

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199-231. https://doi.org/10.1214/ss/1009213726

Brown, W. M., Gedeon, T. D., Groves, D. I., & Barnes, R. G. (2000). Artificial neural networks: A new method for mineral prospectivity mapping. *Australian Journal of Earth Sciences*. https://www.tandfonline.com/doi/abs/10.1046/j.1440-0952.2000.00807.x

Buddenbaum, H., & Steffens, M. (2012). The Effects of Spectral Pretreatments on Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy. *Applied and Environmental Soil Science*, *2012*(1), 274903. https://doi.org/10.1155/2012/274903

Burbano-Orjuela, H. (2016). El suelo y su relación con los servicios ecosistémicos y la seguridad alimentaria. *Revista de Ciencias Agrícolas*, *33*(2), Article 2. https://doi.org/10.22267/rcia.163302.58

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J. T., & Reichstein, M. (2014). Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature*, *514*(7521), 213-217. https://doi.org/10.1038/nature13731

Chang, C.-W., & Laird, D. A. (2002). Near-infrared reflectance spectroscopic analysis of soil c and n. *Soil Science*, *167*(2), 110.

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth*, *89*(B7), 6329-6340. https://doi.org/10.1029/JB089iB07p06329

Clunes, J., Valle, S., Dörner, J., Martínez, O., Pinochet, D., Zúñiga, F., & Blum, W. E. H. (2022). Soil fragility: A concept to ensure a sustainable use of soils. *Ecological Indicators*, *139*, 108969. https://doi.org/10.1016/j.ecolind.2022.108969

Cohen, M. J., Prenger, J. P., & DeBusk, W. F. (2005). Visible-Near Infrared Reflectance Spectroscopy for Rapid, Nondestructive Assessment of Wetland Soil Quality. Journal of Environmental Quality, 34(4), 1422-1434. https://doi.org/10.2134/jeq2004.0353

Comino, F., Aranda, V., García-Ruiz, R., Ayora-Cañada, M. J., & Domínguez-Vidal, A. (2018). Infrared spectroscopy as a tool for the assessment of soil biological quality in agricultural soils under contrasting management practices. *Ecological Indicators*, *87*, 117-126. https://doi.org/10.1016/j.ecolind.2017.12.046

Cotching, W. E. (2018). Organic matter in the agricultural soils of Tasmania, Australia – A review. *Geoderma*, *312*, 170-182. https://doi.org/10.1016/j.geoderma.2017.10.006

Cotrufo, M. F., & Lavallee, J. M. (2022). Chapter One - Soil organic matter formation, persistence, and functioning: A synthesis of current

understanding to inform its conservation and regeneration. En D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 172, pp. 1-66). Academic Press. https://doi.org/10.1016/bs.agron.2021.11.002

Cozzolino, D., & Morón, A. (2003). The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *The Journal of Agricultural Science*, *140*(1), 65-71.

Dai, L., Wang, Z., Zhuo, Z., Ma, Y., Shi, Z., & Chen, S. (2025). Prediction of soil organic carbon fractions in tropical cropland using a regional visible and near-infrared spectral library and machine learning. *Soil and Tillage Research*, *245*, 106297. https://doi.org/10.1016/j.still.2024.106297

Daniel, K. W., Tripathi, N. K., & Honda, K. (2003). Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). *Soil Research*, *41*(1), 47-59. https://doi.org/10.1071/sr02027

Dardenne, P., Sinnaeve, G., & Baeten, V. (2000). Multivariate Calibration and Chemometrics for near Infrared Spectroscopy: Which Method? *Journal of Near Infrared Spectroscopy*, *8*(4), 229-237. https://doi.org/10.1255/jnirs.283

Das, B., Chakraborty, D., Singh, V. K., Das, D., Sahoo, R. N., Aggarwal, P., Murgaokar, D., & Mondal, B. P. (2023). Partial least square regression based machine learning models for soil organic carbon prediction using

visible–near infrared spectroscopy. *Geoderma Regional*, *33*, e00628. https://doi.org/10.1016/j.geodrs.2023.e00628

Di Iorio, E., Napoletano, P., Circelli, L., Memoli, V., Santorufo, L., De Marco, A., & Colombo, C. (2022). Comparison of natural and technogenic soils developed on volcanic ash by Vis-NIR spectroscopy. *CATENA*, *216*, 106369. https://doi.org/10.1016/j.catena.2022.106369

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Casanova Pinto, M., Casanova-Katny, A., Muñoz, C., Boudin, M., Zagal Venegas, E., & Boeckx, P. (2015). Soil carbon storage controlled by interactions between geochemistry and climate. *Nature Geoscience*, *8*(10), 780-783. https://doi.org/10.1038/ngeo2516

Eivazi, F., & Tabatabai, M. A. (1988). Glucosidases and galactosidases in soils. *Soil Biology and Biochemistry*, *20*(5), 601-606. https://doi.org/10.1016/0038-0717(88)90141-1

Ellies, A. (1990). *El rol de la física de suelos en la producción agrícola* (Agro Sur).

Engel, J., Gerretzen, J., Szymańska, E., Jansen, J. J., Downey, G., Blanchet, L., & Buydens, L. M. C. (2013). Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry*, *50*, 96-106. https://doi.org/10.1016/j.trac.2013.04.015

FAO. (2024). *Una Introducción al Análisis de Suelo Utilizando Espectroscopía Visible e Infrarrojo Cercano (vis-NIR) y Espectroscopía de Infrarrojo*

Medio (MIR). https://openknowledge.fao.org/items/c9c82d96-1fe4-4afd-8fd0-02bfa99d3485

Gandariasbeitia, M., Besga, G., Albizu, I., Larregla, S., & Mendarte, S. (2017). Prediction of chemical and biological variables of soil in grazing areas with visible- and near-infrared spectroscopy. *Geoderma*, *305*, 228-235. https://doi.org/10.1016/j.geoderma.2017.05.045

Gobrecht, A., Roger, J.-M., & Bellon-Maurel, V. (2014). Chapter Four - Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation: A Review. En D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 123, pp. 145-175). Academic Press. https://doi.org/10.1016/B978-0-12-420225-2.00004-2

Greenberg, I., Seidel, M., Vohland, M., & Ludwig, B. (2022). Performance of field-scale lab vs in situ visible/near- and mid-infrared spectroscopy for estimation of soil properties. *European Journal of Soil Science*, *73*(1), e13180. https://doi.org/10.1111/ejss.13180

Griffiths, P. R. (2010). Introduction to the Theory and Instrumentation for Vibrational Spectroscopy. En *Handbook of Vibrational Spectroscopy*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470027320.s8935

Herrmann, S., Mayer, J., Michel, K., & Ludwig, B. (2009). Predictive Capacity of Visible-Near Infrared Spectroscopy for Quality Parameter Assessment of Compost. *Journal of Near Infrared Spectroscopy*, *17*(5), 289-301. https://doi.org/10.1255/jnirs.850

Islam, K., Singh, B., & Mcbratney, A. (2003). Simultaneous Estimation of Several Soil Properties by Ultra-Violet, Visible, and Near-Infrared Reflectance Spectroscopy. *Australian Journal of Soil Research - AUST J SOIL RES*, *41*. https://doi.org/10.1071/SR02137

ISO 11465. (1993). Soil quality—Determination of dry matter and water content on a mass basis—Gravimetric method. *International Organization for Standardization-Switzerland*, 3.

Kazeev, K. Sh., Odabashian, M. Yu., Trushkov, A. V., & Kolesnikov, S. I. (2020). Assessment of the Influence of Pyrogenic Factors on the Biological Properties of Chernozems. *Eurasian Soil Science*, *53*(11), 1610-1619. https://doi.org/10.1134/S106422932011006X

Lavallee, J. M., Soong, J. L., & Cotrufo, M. F. (2020). Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. *Global Change Biology*, *26*(1), 261-273. https://doi.org/10.1111/gcb.14859

Li, S., Shi, Z., Chen, S., Ji, W., Zhou, L., Yu, W., & Webster, R. (2015). In Situ Measurements of Organic Carbon in Soil Profiles Using vis-NIR Spectroscopy on the Qinghai–Tibet Plateau. *Environmental Science & Technology*, *49*(8), 4980-4987. https://doi.org/10.1021/es504272x

Li, X., Pan, W., Li, D., Gao, W., Zeng, R., Zheng, G., Cai, K., Zeng, Y., & Jiang, C. (2024). Can fusion of vis-NIR and MIR spectra at three levels improve

the prediction accuracy of soil nutrients? *Geoderma*, *441*, 116754. https://doi.org/10.1016/j.geoderma.2023.116754

Liaw, A., & Wiener. (2002). Classification and Regression by RandomForest. *R News*.

Ludwig, B., Sawallisch, A., Heinze, S., Joergensen, R. G., & Vohland, M. (2015). Usefulness of middle infrared spectroscopy for an estimation of chemical and biological soil properties – Underlying principles and comparison of different software packages. *Soil Biology and Biochemistry*, *86*, 116-125. https://doi.org/10.1016/j.soilbio.2015.03.015

Magdoff, F. (2001). Concept, Components, and Strategies of Soil Health in Agroecosystems. *Journal of Nematology*, *33*(4), 169-172.

Mallah, S., Delsouz Khaki, B., Davatgar, N., Scholten, T., Amirian-Chakan, A., Emadi, M., Kerry, R., Mosavi, A. H., & Taghizadeh-Mehrjardi, R. (2022). Predicting Soil Textural Classes Using Random Forest Models: Learning from Imbalanced Dataset. *Agronomy*, *12*(11), Article 11. https://doi.org/10.3390/agronomy12112613

Mann, C., Lynch, D., Fillmore, S., & Mills, A. (2019). Relationships between field management, soil health, and microbial community composition. *Applied Soil Ecology*, *144*, 12-21. https://doi.org/10.1016/j.apsoil.2019.06.012

McBratney, Alex. B., & Hartemink, A. E. (2024). Define soil. *Soil Security*, *14*, 100135. https://doi.org/10.1016/j.soisec.2024.100135

McClure, W. F. (2003). 204 Years of near Infrared Technology: 1800–2003. *Journal of Near Infrared Spectroscopy*, *11*(6), 487-518. https://doi.org/10.1255/jnirs.399

Metzger, K., Liebisch, F., Herrera, J. M., Guillaume, T., Walder, F., & Bragazza, L. (2024). The use of visible and near-infrared spectroscopy for in-situ characterization of agricultural soil fertility: A proposition of best practice by comparing scanning positions and spectrometers. *Soil Use and Management*, *40*(1), e12952. https://doi.org/10.1111/sum.12952

Montanarella, L., Pennock, D., Mckenzie, S., Alavipanah, S. K., & Zhang, G. L. (2015). *Status of the World's Soil Resources: Main Report*. https://openknowledge.fao.org/server/api/core/bitstreams/6ec24d75-19bd-4f1f-b1c5-5becf50d0871/content

Mouazen, A. M., Kuang, B., De Baerdemaeker, J., & Ramon, H. (2010). Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*, *158*(1), 23-31. https://doi.org/10.1016/j.geoderma.2010.03.001

Nguyen, T. T., Janik, L. J., & Raupach, M. (1991). Diffuse reflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. *Soil Research*, *29*(1), 49-67. https://doi.org/10.1071/sr9910049

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J. A. M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., … Wetterlind, J. (2015). Chapter Four - Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. En D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 132, pp. 139-159). Academic Press. https://doi.org/10.1016/bs.agron.2015.02.002

NRCS, U. (2019). Soil Health Assessment | Natural Resources Conservation Service. *Soil Health. United States Department of Agriculture Natural Resources Conservation Service.* https://www.nrcs.usda.gov/conservation-basics/natural-resource-concerns/soils/soil-health/soil-health-assessment

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional*, *16*, e00198. https://doi.org/10.1016/j.geodrs.2018.e00198

Paul, E. A. (2007). 1—SOIL MICROBIOLOGY, ECOLOGY, AND BIOCHEMISTRY IN PERSPECTIVE. En E. A. Paul (Ed.), *Soil Microbiology, Ecology and Biochemistry (Third Edition)* (pp. 3-24). Academic Press. https://doi.org/10.1016/B978-0-08-047514-1.50005-6

Peng, Y., Ben-Dor, E., Biswas, A., Chabrillat, S., Demattê, J. A. M., Ge, Y., Gholizadeh, A., Gomez, C., Guerrero, C., Herrick, J., Maynard, J. J.,

Mouazen, A. M., Ma, Y., McBratney, A. B., Minasny, B., Ramirez-Lopez, L., Robertson, A. H. J., Viscarra Rossel, R. A., Shi, Z., … Zhang, G. (2025). Spectroscopic solutions for generating new global soil information. *The Innovation*, 100839. https://doi.org/10.1016/j.xinn.2025.100839

Picollo, M., Aceto, M., & Vitorino, T. (2019). UV-Vis spectroscopy. *Physical Sciences Reviews*, *4*(4). https://doi.org/10.1515/psr-2018-0008

Poeplau, C., & Don, A. (2013). Sensitivity of soil organic carbon stocks and fractions to different land-use changes across Europe. *Geoderma*, *192*, 189-201. https://doi.org/10.1016/j.geoderma.2012.08.003

Poppiel, R. R., Paiva, A. F. da S., & Demattê, J. A. M. (2022). Bridging the gap between soil spectroscopy and traditional laboratory: Insights for routine implementation. *Geoderma*, *425*, 116029. https://doi.org/10.1016/j.geoderma.2022.116029

Preethi, B., Poorniammal, R., Balachandar, D., Karthikeyan, S., Chendrayan, K., Bhattacharyya, P., & Adhya, T. K. (2013). Long-term organic nutrient managements foster the biological properties and carbon sequestering capability of a wetland rice soil. *Archives of Agronomy and Soil Science*, *59*(12), 1607-1624. https://doi.org/10.1080/03650340.2012.755260

Ramifehiarivo, N., Barthès, B. G., Cambou, A., Chapuis-Lardy, L., Chevallier, T., Albrecht, A., & Razafimbelo, T. (2023). Comparison of near and mid-infrared reflectance spectroscopy for the estimation of soil organic carbon

fractions in Madagascar agricultural soils. *Geoderma Regional*, *33*, e00638. https://doi.org/10.1016/j.geodrs.2023.e00638

Reyna, L., Dube, F., Barrera, J. A., & Zagal, E. (2017). Potential Model Overfitting in Predicting Soil Carbon Content by Visible and Near-Infrared Spectroscopy. *Applied Sciences*, *7*(7), Article 7. https://doi.org/10.3390/app7070708

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), 1201-1222. https://doi.org/10.1016/j.trac.2009.07.007

Rodríguez P, W., García R, P. A., & Fajardo O, A. (2016). Aplicaciones de técnicas espectroscópicas para el análisis de suelos. *Revista Facultad de Ciencias Básicas*, *12*(2), Article 2. https://doi.org/10.18359/rfcb.2030

Rodriguez-Galiano, V. F., Chica-Olmo, M., & Chica-Rivas, M. (2014). Predictive modelling of gold potential with the integration of multisource information based on random forest: A case study on the Rodalquilar area, Southern Spain. *International Journal of Geographical Information Science*. https://www.tandfonline.com/doi/full/10.1080/13658816.2014.885527

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of*

Photogrammetry and Remote Sensing, *67*, 93-104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rossel, R. A. V., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, *158*(1), 46-54. https://doi.org/10.1016/j.geoderma.2009.12.025

Sankey, J. B., Brown, D. J., Bernard, M. L., & Lawrence, R. L. (2008). Comparing local vs. Global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. *Geoderma*, *148*(2), 149-158. https://doi.org/10.1016/j.geoderma.2008.09.019

Schnürer, J., & Rosswall, T. (1982). Fluorescein Diacetate Hydrolysis as a Measure of Total Microbial Activity in Soil and Litter. *Applied and Environmental Microbiology*, *43*(6), 1256-1261. https://doi.org/10.1128/aem.43.6.1256-1261.1982

Sepúlveda, M. de los Á., 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. https://doi.org/10.1016/j.geodrs.2021.e00397

Sergeevna Kozun, Y., Shagidullovich Kazeev, K., & Ilich Kolesnikov, S. (2022). Climatic gradients of biological properties of zonal soils of natural lands.

Geoderma, *425*, 116031. https://doi.org/10.1016/j.geoderma.2022.116031

Shao, J. (1993). Linear Model Selection by Cross-validation. *Journal of the American Statistical Association*, *88*(422), 486-494. https://doi.org/10.1080/01621459.1993.10476299

Shaw, G. A., & Burke, H. K. (2003). *Spectral Imaging for Remote Sensing*. *14*(1).

Sherrod, L. A., Reeder, J. D., Hunter, W., & Ahuja, L. R. (2012). Rapid and Cost-Effective Method for Soil Carbon Mineralization in Static Laboratory Incubations. *Communications in Soil Science and Plant Analysis*, *43*(6), 958-972. https://doi.org/10.1080/00103624.2012.653031

Skjemstad, J., Spouncer, L., Cowie, B., & Swift, R. (2004). Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. *Australian Journal of Soil Research*, *42*. https://doi.org/10.1071/SR03013

Song, J., Shi, X., Wang, H., Lv, X., Zhang, W., Wang, J., Li, T., & Li, W. (2024). Improving Soil Quality Index Prediction by Fusion of Vis-NIR and pXRF spectral data. *Geoderma*, *447*, 116938. https://doi.org/10.1016/j.geoderma.2024.116938

Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., 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. https://doi.org/10.1080/05704928.2013.811081

Staff, S. S. (1999). *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys.* U.S. Department of Agriculture, Natural Resources Conservation Service.

Stark, C., Condron, L. M., Stewart, A., Di, H. J., & O'Callaghan, M. (2007). Influence of organic and mineral amendments on microbial soil properties and processes. *Applied Soil Ecology*, *35*(1), 79-93. https://doi.org/10.1016/j.apsoil.2006.05.001

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Chapter Five—Visible and Near Infrared Spectroscopy in Soil Science. En D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 107, pp. 163-215). Academic Press. https://doi.org/10.1016/S0065-2113(10)07005-7

Tabatabai, M. A. (1982). *Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties* (Page, A.L., Miller, E.M., Keeney, D.R. (Eds)). Soil Science Society of America, Inc, Madison.

Tan, K., Wang, H., Chen, L., Du, Q., Du, P., & Pan, C. (2020). Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. *Journal of Hazardous Materials*, *382*, 120987. https://doi.org/10.1016/j.jhazmat.2019.120987

Telo da Gama, J. (2023). The Role of Soils in Sustainability, Climate Change, and Ecosystem Services: Challenges and Opportunities. *Ecologies*, *4*(3), Article 3. https://doi.org/10.3390/ecologies4030036

Terhoeven-Urselmans, T., Schmidt, H., Georg Joergensen, R., & Ludwig, B. (2008a). Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: Importance of sample pre-treatment. *Soil Biology and Biochemistry*, *40*(5), 1178-1188. https://doi.org/10.1016/j.soilbio.2007.12.011

Terhoeven-Urselmans, T., Schmidt, H., Georg Joergensen, R., & Ludwig, B. (2008b). Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: Importance of sample pre-treatment. *Soil Biology and Biochemistry*, *40*(5), 1178-1188. https://doi.org/10.1016/j.soilbio.2007.12.011

Valle, S. R., & Carrasco, J. (2018). Soil quality indicator selection in Chilean volcanic soils formed under temperate and humid conditions. *CATENA*, *162*, 386-395. https://doi.org/10.1016/j.catena.2017.10.024

Vašát, R., Kodešová, R., Borůvka, L., Jakšík, O., Klement, A., & Brodský, L. (2017). Combining reflectance spectroscopy and the digital elevation model for soil oxidizable carbon estimation. *Geoderma*, *303*, 133-142. https://doi.org/10.1016/j.geoderma.2017.05.018

Viscarra Rossel, R. A., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, *158*(1), 46-54. https://doi.org/10.1016/j.geoderma.2009.12.025

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Ge, Y., Gomez, C., Guerrero, C., Peng, Y., Ramirez-Lopez, L., Shi, Z., Stenberg, B., Webster, R., Winowiecki, L., & Shen, Z. (2022). Diffuse reflectance spectroscopy for estimating soil properties: A technology for the 21st century. *European Journal of Soil Science*, *73*(4), e13271. https://doi.org/10.1111/ejss.13271

Viscarra Rossel, R. A., & Hicks, W. S. (2015). Soil organic carbon and its fractions estimated by visible–near infrared transfer functions. *European Journal of Soil Science*, *66*(3), 438-450. https://doi.org/10.1111/ejss.12237

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, *131*(1), 59-75. https://doi.org/10.1016/j.geoderma.2005.03.007

Vohland, M., Ludwig, M., Thiele-Bruhn, S., & Ludwig, B. (2014). Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*, *223-225*, 88-96. https://doi.org/10.1016/j.geoderma.2014.01.013

Walsh, E., & McDonnell, K. P. (2012). The influence of added organic matter on soil physical, chemical, and biological properties: A small-scale and short-time experiment using straw. *Archives of Agronomy and Soil Science*, *58*(sup1), S201-S205. https://doi.org/10.1080/03650340.2012.697999

Wang, X., Zhang, M.-W., Guo, Q., Yang, H.-L., Wang, H.-L., & Sun, X.-L. (2023). Estimation of soil organic matter by in situ Vis-NIR spectroscopy using an automatically optimized hybrid model of convolutional neural network and long short-term memory network. *Computers and Electronics in Agriculture*, *214*, 108350. https://doi.org/10.1016/j.compag.2023.108350

Wold, S., Antti, H., Lindgren, F., & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, *44*(1), 175-185. https://doi.org/10.1016/S0169-7439(98)00109-9

Wold, S., Martens, H., & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. En B. Kågström & A. Ruhe (Eds.), *Matrix Pencils* (pp. 286-293). Springer. https://doi.org/10.1007/BFb0062108

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109-130. https://doi.org/10.1016/S0169-7439(01)00155-1

Xu, S., Wang, M., Shi, X., Yu, Q., & Zhang, Z. (2021). Integrating hyperspectral imaging with machine learning techniques for the high-resolution mapping of soil nitrogen fractions in soil profiles. *Science of The Total Environment*, *754*, 142135. https://doi.org/10.1016/j.scitotenv.2020.142135

Yang, J., Wang, X., Wang, R., & Wang, H. (2020). Combination of Convolutional Neural Networks and Recurrent Neural Networks for predicting soil properties using Vis–NIR spectroscopy. *Geoderma*, *380*, 114616. https://doi.org/10.1016/j.geoderma.2020.114616

Yang, X. M., Xie, H. T., Drury, C. F., Reynolds, W. D., Yang, J. Y., & Zhang, X. D. (2012a). Determination of organic carbon and nitrogen in particulate organic matter and particle size fractions of Brookston clay loam soil using infrared spectroscopy. *European Journal of Soil Science*, *63*(2), 177-188. https://doi.org/10.1111/j.1365-2389.2011.01421.x

Zabaloy, M. C. (2021). Una sola salud: La salud del suelo y su vínculo con la salud humana. *Revista Argentina de Microbiología*, *53*(4), 275-276. https://doi.org/10.1016/j.ram.2021.11.001

Zhang, J., Li, Y., Jia, J., Liao, W., Amsili, J. P., Schneider, R. L., van Es, H. M., Li, Y., & Zhang, J. (2023). Applicability of soil health assessment for wheat-maize cropping systems in smallholders' farmlands. *Agriculture, Ecosystems & Environment*, *353*, 108558. https://doi.org/10.1016/j.agee.2023.108558

Zhang, T., Li, Y., & Wang, M. (2024). Prediction of soil organic carbon and total nitrogen affected by mine using Vis–NIR spectroscopy coupled with machine learning algorithms in calcareous soils. *Scientific Reports*, *14*(1), 28014. https://doi.org/10.1038/s41598-024-73761-6

Zhong, L., Guo, X., Xu, Z., & Ding, M. (2021). Soil properties: Their prediction and feature extraction from the LUCAS spectral library using deep convolutional neural networks. *Geoderma*, *402*, 115366. https://doi.org/10.1016/j.geoderma.2021.115366

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K. M., Arcenegui, V., & Mataix-Beneyto, J. (2008). Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biology and Biochemistry*, *40*(7), 1923-1930. https://doi.org/10.1016/j.soilbio.2008.04.003

Tables

Table 1 Predicted biological properties in the Mid Infrared (MIR) and Visible and near infrared (Vis-NIR) range.

Biological properties	Validation type	Vis-NIR/NIR	MIR	Reference
Basal soil respiration	Cross validation	R^2 = 0.92	R^2 =0.63	(Zornoza et al., 2008); (Ludwig *et al.,* 2015)
Microbial biomass-C	Cross validation	R^2 = 0,61	R^2 = 0.78	(Ludwig *et al.,* 2015)
Microbial biomass-N	External validation		R^2 = 0.73	(Di Iorio *et al.,* 2022)
Microbial biomass-P	Cross validation	R^2 = 0,81		(Terhoeven-Urselmans et al., 2008)
Nitrogen mineralisation rate	Cross validation	R^2 = 0.54		(Terhoeven-Urselmans et al., 2008)
Enzime/ Dehydrogenase	Cross validation	R^2 = 0.23		(Comino *et al.,* 2018)
β-glucosidase	Cross validation		R^2= 0.93	(Comino *et al.,* 2018)
Particulate Organic Matter-C	Cross validation		R^2 = 0,92	(X. M. Yang et al., 2012)
Particulate Organic Matter-C	External validation	R^2 =71		(Dai et al., 2025)

Table 2 Physical and chemical properties of soil samples.

Variable	Mean	S. D	Mín	Máx
C	0.22	0.21	1.00E-03	0.72
N	3.01	2.38	0.44	8.79
pH	5.77	0.45	4.7	7.18
MO	4.94	3.98	0.68	14.65
N_NO3	32.83	29.87	3.1	134.5
N_NH4	6.2	6.03	2.3	49.9
N dis	39.03	29.96	8.5	137.4
P_Olsen	17.01	18.52	0.4	85.2
K dis	204.83	121.45	36.9	678.9
K int	0.52	0.31	0.09	1.74
Ca dis	6	3.22	0.8	18.25
Sand	32.13	17.31	2.9	82.7
Silt	38.62	12.85	9	76.3
Clay	29.22	11.68	7.3	56.2
DA	1.29	0.33	0.64	1.96

C: Carbon, N: Nitrogen, pH: Hydrogen potential, MO: Organic matter, N_NO3: Nitrogen nitrate, N_NH4: Ammonium nitrogen, N_dis: Available nitrogen, P_Olsen: Available phosphorus, K_dis: Available potassium, K_int: Exchangeable potassium, DA: Apparent density, S.D: Standard deviation, Min: minimum, Max: Maximum, n=70

Table 3 Chemical analysis results for soil respiration, β-glucosidase activity, microbial biomass (FDA), and particulate organic matter-C.

Variables	Units	S. D	Mín.	Máx.	Mean
Soil Respiration	mg gss^{-1}	0.31	0.35	1.88	0.85
Microbial Biomass (FDA)	µF gss^{-1}	15.83	7.33	82.07	32.52
Enzyme/β-Glucosidase	µm PNF gss^{-1}	0.47	0.14	2.35	1.01
Particulate Organic Matter-C	%TC	2.35	0.05	10.53	2.73

mg gss^{-1}: Milligrams of gas (CO_2) per gram of dry soil per unit of time FDA: Fluorescein diacetate hydrolysis, µF gss^{-1}: Microgram of fluorescein per gram of dry soil, µm PNF gss^{-1}: Microgram of p-nitrophenol per gram of dry soil, %TC: Total carbon, S. D. Standard deviation, Min: minimum, Max: maximum. n=70

Table 4 Partial least squares regression (PLSR) models for predicting soil respiration, β-glucosidase activity, microbial biomass (FDA), and particulate organic matter-C.

Variable	Transformation	Cumulative of variance	R^2 Cal	R^2 Val	RMSECV	# of factors
Soil respiration	OSC#1+MSC+Normalize	95.0	0.99	0.99	0.05	5
Soil respiration	MSC+2°Deriv+Baseline Correction	90.8	0.83	0.55	0.207	6
Enzyme/β-glucosidase activity	OSC#1+MSC+Normalize	94.7	0.98	0.98	0.101	5
Enzyme/β-glucosidase activity	Normalize+Log10+2°Deriv+Smooth	79.1	0.89	0.63	0.21	6
Microbial biomass (FDA)	OSC#1+Log10+Baseline Correction	92.9	0.95	0.92	10.108	5 *
Microbial biomass (FDA)	Log10+MSC+Baseline correction	97.5	0.62	0.41	10.205	6
Particulate organic matter-C	Log10+Normalize	99.6	0.50	0.30	1.587	5
Particulate organic matter-C	OSC#1+Baseline Correction	93.8	0.95	0.95	1.486	4

RMSEcv: root mean square error of cross-validation, rval: r of validation, rcal: r of calibration, FDA: Fluorescein diacetate hydrolysis, OSC: Orthogonal signal correction, MSC: Multiplicative signal correction, 2° deriv: Second, derivative, #1= Component, *Microbial Biomass (FDA)= outlier= 3,15,16

Table 5 Random Forest (RF) models for predicting soil respiration, β-glucosidase activity, microbial biomass (FDA), and particulate organic matter-C.

Variable	RPD	r Val	RMSEP
Soil respiration	1.24	0.35	0.223
Enzyme/β-glucosidase	1.34	0.45	0.274
Microbial biomass (FDA)	1.30	0.41	8.01
Particulate organic matter-C	2.08	0.77	1.128

mg gss^{-1}: Milligrams of gas (CO_2) per gram of dry soil per unit of time FDA: Fluorescein diacetate hydrolysis, µF gss^{-1}: Microgram of fluorescein per gram of dry soil, µm PNF gss^{-1}: Microgram of p-nitrophenol per gram of dry soil, %TC: Total carbon, RDP: ratio of standard deviation, RMSE: root mean square root

Figures

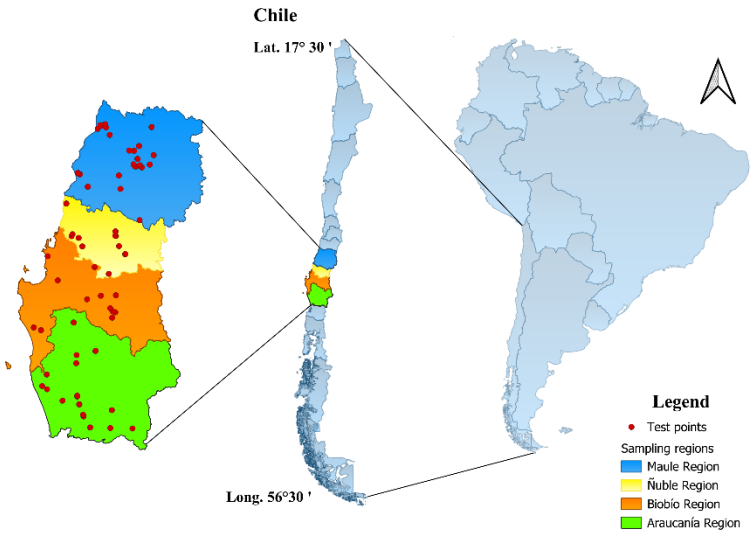

Figure 1 Distribution of study sites across the Maule, Ñuble, Biobío, and Araucanía regions. Red points indicate locations where soil samples were collected, situated in the Andean foothills, central and coastal depression of Chile.

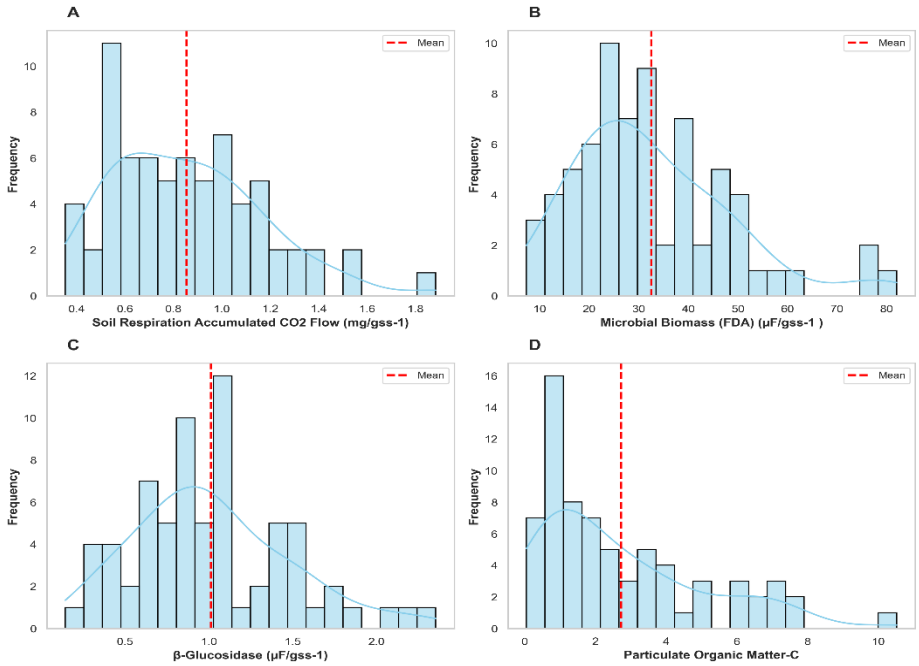

Figure 2 Data distribution of traits: A) soil respiration, B) microbial biomass (FDA), C) β-glucosidase activity, and D) particulate organic matter-C.

Equations

Equation 1 Mean square error of cross-validation.

$$RMSECV = \sqrt{\sum_{i=1}^{Y_0} \frac{(y_{Y0} - \hat{y}_{Y0})^2}{n}}$$

Equation 2 Root Mean Square Prediction Error.

$$RMSEP = \sqrt{\frac{\sum_{i-1}^{i=n}(x_i - y_i - bias)^2}{n-1}}$$

Equation 3 Coefficient of determination.

$$R^2 = 1\frac{\sum_{i=1}^{Y_0}(y_{Y0} - \hat{y}_{Y0})^2}{\sum_{i=1}^{Y_0}(y_{Y0} - \bar{y}_{Y0})^2}$$

Equation 4 Ratio of Performance to Deviation.

$$RDP = \frac{SD}{RMSEP}$$

Conclusiones generales

La espectroscopía de suelos, especialmente en el rango visible e infrarrojo cercano (Vis-NIR), representa un avance científico significativo para la caracterización bioquímica edáfica, permitiendo la predicción rápida y no destructiva de parámetros clave como la actividad enzimática, la biomasa microbiana y el carbono orgánico. Su capacidad para generar datos digitalizados en tiempo reducido responde a la necesidad crítica de monitorear la salud del suelo ante escenarios de cambio climático y demandas agrícolas intensivas. Sin embargo, su implementación operativa enfrenta desafíos inherentes a la heterogeneidad espacial y temporal de los suelos, evidenciados por la brecha entre los altos desempeños en condiciones controladas y la disminución de precisión en validaciones externas, particularmente en ambientes geoclimáticamente diversos.

La superación de estas limitaciones requiere estrategias integradas, como la ampliación del rango espectral (ejemplo, MIR) y la utilización de bibliotecas espectrales representativas, el desarrollo de modelos locales basados en agrupamientos pedogenéticos (ej., suelos volcánicos) y la adopción de preprocesamientos avanzados que minimicen el ruido ambiental. Paralelamente, iniciativas globales lideradas por consorcios científicos están trabajando en la estandarización de protocolos y la promoción de colaboraciones interinstitucionales, aspectos fundamentales para consolidar esta tecnología

como herramienta accesible y confiable. En este marco, la espectroscopía de suelos se proyecta no solo como un método analítico innovador, sino como un pilar para la gestión sostenible de ecosistemas, siempre que se aborden de manera sistemática los desafíos técnicos, metodológicos y escalares que aún persisten.

Supplementary material

Table S1 General information of the sites, ID, UTM _N, UTM _E, parent matter, geomorphology, physiographic position.

ID	UTM _N	UTM _E	Parent Material	Geomorphology	Physiographic Position
1	5647302.11	716739,3	Continental sedimentary	Central plain with moraines and cones	Plain
2	5644301.48	748359,13	Continental sedimentary	Central plain with moraines and cones	Plain
3	5642970.48	265763,25	Continental sedimentary	Active volcanic mountain range	High slope
4	5606158.45	643716,20	Metamorphic	Coastal mountain range	High slope
5	5692118.11	676271,66	Metamorphic	Fluvial or alluvial sedimentation plains	Plateau
6	5711597.56	653152,38	Metamorphic	Moraine barrier lakes	Half slope
7	5698433.81	699795,52	Continental sedimentary	Central plain with moraines and cones	Plain
8	5698433.81	699795,52	Continental sedimentary	Central plain with moraines and cones	Plain
9	5749405.07	701587,62	Volcanic	Central plain with moraines and cones	Valley bottom
10	5762077.84	702992,43	Volcanic	Central plain with moraines and cones	Half slope
11	5813234.4	701736,74	Volcanic	Central plain with moraines and cones	Valley bottom
12	5813234.4	701736,74	Volcanic	Central plain with moraines and cones	Plain
13	5766650.59	733027,74	Volcanic	Central plain with moraines and cones	Half slope
14	5766650.6	733027,75	Volcanic	Central plain with moraines and cones	Half slope
15	5734736.29	654216,18	Metamorphic	Coastal mountain range	Half slope
16	5931423.47	722834,31	Plutonic	Central fluvio-glacio-volcanic plain	High slope
17	No data	No data	Metamorphic	Marine or fluviomarine plain	Half slope
18	5945114.18	717932,92	Plutonic	Marginal granite basins	Half slope
19	5852013.35	239737,56	Continental sedimentary volcano	Pre-mountains	Low slope
20	5816949.67	762508,10	Continental sedimentary	Central plain with moraines and cones	Low slope
21	5852427.49	746913,11	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain

22	239013	5952174,35	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
23	239013	5952174,35	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
24	5928676.22	756176,66	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
25	5945053.5	239822,38	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
26	244632.87	5929137,85	Continental sedimentary	Central fluvio-glacio-volcanic plain	Half slope
27	238802.39	5945395,26	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
28	708511.31	5951534,85	Plutonic	Marginal granite basins	Half slope
29	706921.44	5948441,83	Plutonic	Coastal mountain range	Crest
30	5898011.64	740074,29	Continental sedimentary volcano	Central fluvio-glacio-volcanic plain	Plain
31	5825320	238791,93	Continental sedimentary	Pre-mountains	Plain
32	5826610.49	764191,43	Continental sedimentary	Pre-mountains	Plain
33	5804564.12	649757,12	Metamorphic	Coastal mountain range	Half slope
34	No data	No data	Continental sedimentary	Precordillera	Plain
35	5919548	667512	Continental sedimentary	Marine or fluviomarine plain	Plain
36	5880746	680970	Continental sedimentary	Coastal mountain range	Half slope
37	5809515	638717,00	Marine and continental sedimentary	Coastal mountain range	High slope
38	5831976	760318,00	Continental sedimentary volcanic	Pre-mountains	Half slope
39	6057403	267378	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
40	605428	269593	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
41	6054280	269593	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
42	6052634	280502	Continental sedimentary	Central fluvio-glacio-volcanic plain	Low slope
43	6055907	277334	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
44	6055906	277237	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
45	6054260	271802	Continental sedimentary	Central fluvio-glacio-volcanic plain	High slope
46	6071162	299020	Volcanic	Pre-mountains	High slope

47	6130741	29846	Continental sedimentary	Coastal mountain range	Half slope
48	6115836	295854	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
49	6115840	295846	Continental sedimentary	Central fluvio-glacio-volcanic plain	Flat sector
50	6078510	268332	Volcanic	Central fluvio-glacio-volcanic plain	Plain
51	6086162	276020	Volcanic	Central fluvio-glacio-volcanic plain	Plain
52	6039936	244371	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
53	6024336	737189	Continental sedimentary	Marginal granite basins	Half slope
54	6056940	292924	Volcanic	Pre-mountains	Valley bottom
55	6046447	723080	Marine and transitional sedimentary	Marine or fluviomarine plain	Low slope
56	6044515	726104	Marine and transitional sedimentary	Marine or fluviomarine plain	Low slope
57	6046844	723181	Marine and transitional sedimentary	Marine or fluviomarine plain	Low slope
58	6065861	273737	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
59	6024346	737207	Continental sedimentary	Marginal granite basins	Half slope
60	6103515	230025	Plutonic	Coastal mountain range	Half slope
61	6115440	771360	Continental sedimentary	Coastal mountain range	Low slope
62	6118250	765564	Continental sedimentary	Marine or fluviomarine plain	Low slope
63	6120640	770088	Metamorphic	Coastal mountain range	Low slope
64	6118227	765548	Continental sedimentary	Marine or fluviomarine plain	Low slope
65	6119492	762775	Continental sedimentary	Marine or fluviomarine plain	Low slope
66	6114000	758197	Continental sedimentary	Marine or fluviomarine plain	Terrace
67	5969701.28	276933,23	Volcano-sedimentary	Pre-mountains	Depression
68	6056940	292924	Volcanic	Pre-mountains	Valley bottom
69	6065851	273687	Continental sedimentary	Central fluvio-glacio-volcanic plain	Plain
70	6018905	246907	Volcanic	Central fluvio-glacio-volcanic plain	Plain