



UNIVERSIDAD AUTÓNOMA CHAPINGO

DEPARTAMENTO DE IRRIGACIÓN E INGENIERÍA MECÁNICA AGRÍCOLA

POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

Análisis nutricional de suelos con Aprendizaje Automático y Espectroscopía Vis-NIR

TESIS

QUE COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE:

DOCTOR EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA



APROBADA



P R E S E N T A:

REYES RIVERA ALEJANDRO ERIC

BAJO LA SUPERVISIÓN DE:

DR. GILBERTO DE JESÚS LÓPEZ CANTEÑS

Chapingo, Estado de México, noviembre de 2023.

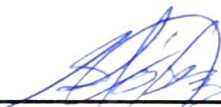


**Análisis nutricional de suelos con Aprendizaje Automático y
Espectroscopía Vis-NIR**

Tesis realizada por **Reyes Rivera Alejandro Eric** bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTOR EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

DIRECTOR:



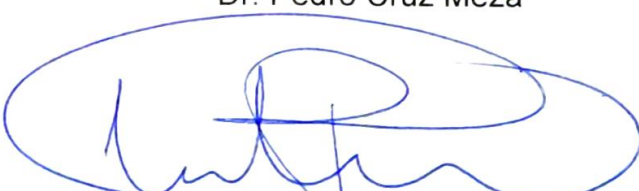
Dr. Gilberto de Jesús López Canteñs

ASESOR:



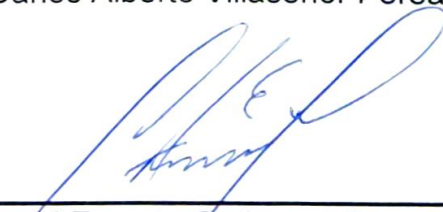
Dr. Pedro Cruz Meza

ASESOR:



Dr. Carlos Alberto Villasenor Perea

LECTOR
EXTERNO:



Dr. Ronald Ernesto Ontiveros Capurata

TABLA DE CONTENIDO

LISTA DE FIGURAS	IV
LISTA DE CUADROS	V
RESUMEN GENERAL.....	1
GENERAL ABSTRACT.....	2
CAPITULO 1: INTRODUCCIÓN GENERAL	3
CAPITULO 2: MARCO TEÓRICO	6
2.1 MACRONUTRIENTES DEL SUELO	6
2.2 DETECCIÓN DE PROPIEDADES NUTRICIONALES EN SUELOS	8
2.3 ESPECTROSCOPÍA	9
2.3.1 Generalidades	10
2.3.2 Principios	11
2.3.3 Espectroscopía para suelos.....	17
2.4 APRENDIZAJE AUTOMATICO PARA ANALISIS ESPECTRAL EN SUELOS ...	25
2.4.1 Aprendizaje automático, generalidades	25
2.4.2 Conceptos básicos en aprendizaje automático	26
2.4.3 Subcampos de aprendizaje automático para suelos	32
2.4.4 Desarrollo de modelos espectrales con aprendizaje automático para predicción de nutrientes de suelo	37
CAPITULO 3. EVALUACIÓN DE TRES MÉTODOS DE SELECCIÓN DE VARIABLES PARA PREDICCIÓN DE NITRÓGENO TOTAL Y PH DEL SUELO USANDO DATOS ESPECTRALES VIS-NIR.....	67
3.1 INTRODUCCIÓN	67
3.2 MATERIALES Y MÉTODOS.....	70
3.2.1 Librería Espectral LUCAS	70
3.2.2 Pre-Procesamiento de los Datos Espectrales	70
3.2.3 Regresión PLS con el Espectro Completo FS-PLS	71
3.2.4 Métodos de selección de variables espectrales	71
3.2.5 Evaluación de los Modelos de Predicción.	73
3.3 RESULTADOS Y DISCUSIÓN	74
3.3.1 TN y pH del suelo en la librería LUCAS	74
3.3.3 Desempeño de los métodos GA-PLS, BFFA-PLS, CARS-PLS y FS-PLS	76
3.3.4 Longitudes de onda seleccionadas por los MVS.....	81
3.4 CONCLUSIONES.....	84

CAPITULO 4. EVALUACIÓN DE ALGORITMOS DE CALIBRACIÓN PARA PREDICCIÓN DE NITRÓGENO DEL SUELO MEDIANTE ESPECTROSCOPIA NIR Y AUMENTO DE DATOS CON REDES GENERATIVAS ADVERSARIAS.	97
4.1. INTRODUCCIÓN	97
4.2. MATERIALES Y METODOS	100
4.2.1 Área de estudio, muestreo y análisis nutricional	100
4.2.2 Análisis espectral.....	101
4.2.3 Desarrollo de los modelos espectrales.....	102
4.2.4. Evaluación	107
4.3. RESULTADOS Y DISCUSIÓN	108
4.3.1 Descripción estadística de los datos de nitrógeno reales y artificiales	108
4.3.2. Análisis de los espectros reales y artificiales.....	110
4.3.3. Impacto del aumento de datos en los algoritmos de regresión.....	112
4.3.4 Comparación del desempeño de los modelos de calibración.....	116
4.4. CONCLUSIONES.....	119

LISTA DE FIGURAS

Figura 2.1 El suelo como sistema trifásico.	7
Figura 2.2 Procesos de absorción y emisión en la interacción materia-radiación.	15
Figura 2.3 Espectro electromagnético y regiones importantes.	17
Figura 2.4. Efectos de dispersión en medios no difusos (A) y difusos (B y C).	19
Figura 2.5. Diagrama de un Espectroscopio FT-IR.	20
Figura 2.6. Diagramas de los métodos de medición de espectros IR de suelos	23
Figura 2.7. Método de análisis espectral de suelos con ML.	40
Figura 2.8 Muestreo en zigzag (A), estratificado(B), cinco de oros(C).	41
Figura 2.9. Muestra estándar de calibración(A), muestra colocada en el espectrómetro (B) y medición de la respuesta espectral(C).	45
Figura 3.1. Espectros de origen promedio de cada clase textural y su segunda derivada SDR.	75
Figura 3.2. Observaciones y predicciones de los mejores modelos PLS con cada MSV. Secciones a, b y c; TN en T4, T2 y T6 con GA, BFFA y CARS respectivamente. Secciones	78

d, e y f; pH en T5, T3 y T6 con GA, BFFA y CARS respectivamente. Se indica el coeficiente R^2 de predicción y el RMSEcv (LOO).	
Figura 3.3. Desempeño de los modelos en función del RPD para la predicción de TN (sección a) y pH (sección b) para cada subconjunto de datos.	80
Figura 3.4. Regiones espectrales seleccionadas (ennegrecidas) por GA, CARS y BFFA para TN (secciones a, c y e respectivamente) y para pH del suelo (secciones b, d y f).	82
Figura 4.1. Campo Experimental “La Xerona”, Chapingo, Texcoco Estado de México.	93
Figura 4.2. Diagrama de los modelos de regresión con aumento de datos.	94
Figura 4.3. Diagrama de caja de las distribuciones del nitrógeno aprovechable (real) y el generado artificialmente.	101
Figura 4.4. Espectros originales(a), Primera derivada de la reflectancia (b).	102
Figura 4.5. Diagramas comparativos de los dos primeros componentes principales de la distribución espectral (a) y comparación grafica de la reflectancia media y su desviación estándar (b) para los conjuntos de datos reales y artificiales.	104
Figura 4.6. Efecto del aumento de datos en los parámetros de predicción de los modelos de regresión PLS(a), AE (b), SVM(c) y BA (d).	106
Figura 4.7. Mejores ajustes de los modelos con aumento de datos; (a) MVS en el quinto nivel, (b) BA en el sexto, (c) AE en el séptimo, (d) PLS en el segundo.	110

LISTA DE CUADROS

Cuadro 2.1. Estados vibratorios en función de la región espectral.	16
Cuadro 2.2. Espectrómetros IR para propiedades del suelo.	25
Cuadro 2.3. Formas de aprendizaje en ML	30
Cuadro 2.4. Métricas de evaluación de modelos de regresión en ML.	33
Cuadro 2.5. Subcampos del Aprendizaje automático.	38
Cuadro 2.6. Pretratamientos para corrección de espectros de suelo.	47
Cuadro 2.7 Métodos de selección de variables para espectroscopia de suelos.	51

Cuadro 2.8. Métodos de regresión para detección de NPK en suelos.	54
Cuadro 3.1. Estadísticas de TN y pH del suelo en la librería LUCAS.	74
Cuadro 3.2. Parámetros de predicción de TN y pH del suelo con modelos de espectro completo FS-PLS en cada clase de suelo.	76
Cuadro 3.3. Parámetros de predicción de TN y pH con modelos GS-PLS, BFFA-PLS y CARS-PLS en cada tipo de suelo.	77
Cuadro 4.1. Estadística descriptiva de los datos de nitrógeno aprovechable (NA) real y artificial.	101
Cuadro 4.2. Resultados del análisis de componentes principales.	103
Cuadro 4.3. Comparación del desempeño de los modelos de calibración a través del aumento gradual de daos de entrenamiento.	107

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT).

A la Universidad Autónoma Chapingo.

Al Posgrado de Ingeniería Agrícola y Uso Integral del Agua.

A mi comité de tesis; agradezco profundamente al **Dr. Gilberto de Jesús López Canteñs** por la orientación y los conocimientos que me ha transmitido a través de estos años. De igual forma, agradezco al **Dr. Pedro Cruz Meza** y al **Dr. Noel Chávez Aguilera** por el apoyo brindado y por disponer parte de su tiempo para culminar este proyecto. Así mismo, un agradecimiento especial al Dr. Ronald Ernesto Ontiveros Capurata y al Dr. Carlos Alberto Villaseñor Perea por fungir como lector externo y miembro del jurado respectivamente para este trabajo de investigación.

DATOS BIOGRÁFICOS

Alejandro Eric Reyes Rivera, autor principal del presente trabajo, originario del estado de México, es egresado del departamento de Ingeniería Mecánica Agrícola (generación 2014) de la Universidad Autónoma Chapingo (UACH). Así mismo, realizó sus estudios de maestría en la misma casa de estudios a través del posgrado en Ingeniería Agrícola y Uso Integral del Agua (generación 2017), donde tiene estudios enfocados a la investigación de propiedades físico-mecánicas de frutos de nuez de castilla y a la aplicación de visión artificial para clasificación e identificación de calidad.

Se ha desarrollado como ingeniero de proyectos para monitoreo a distancia y automatización en empresas como Agropec Star en el estado de Jalisco. Así mismo ha prestado servicios de automatización y monitoreo a productores de hongos y flores de Crisantemo en el municipio de Texcoco de manera independiente. En el campo de la docencia, ha participado como apoyo en la materia de programación del Departamento de Ingeniería Mecánica Agrícola de la UACH e impartiendo las materias de electrónica básica, electrónica digital e hidráulica en el Centro Educativo Grupo CEDVA de Texcoco.

En 2019 comenzó sus estudios doctorales enfocados al estudio y aplicación de Inteligencia Artificial a problemas de teledetección en el campo de la agricultura de precisión, centrándose en el análisis y evaluación de algoritmos de aprendizaje automático y en el uso de técnicas innovadoras como la espectroscopia de suelos.

RESUMEN GENERAL

Análisis nutricional de suelos con Aprendizaje Automático y Espectroscopía Vis-NIR¹

El presente trabajo de tesis se centra en la aplicación de técnicas de Aprendizaje Automático (ML por Machine Learning) y Espectroscopia visible y de infrarrojo cercano (Vis-NIR) para mejorar el análisis nutricional de suelos en un contexto de la agricultura de precisión. El objetivo principal es estudiar y aportar en la mejora de la precisión y eficiencia en la predicción de nutrientes del suelo, en particular el nitrógeno. Se abordan dos enfoques esenciales en la tesis: la selección de variables y la calibración de algoritmos ML. Por un lado, se explora la efectividad de tres métodos de selección de longitudes de onda claves en la predicción de nitrógeno total y pH del suelo, mediante el uso de datos espectrales Vis-NIR y métodos de búsqueda y optimización, incluyendo algoritmos genéticos y bio-inspirados.

Por otro lado, se profundiza en la evaluación de métodos de calibración para la predicción de nitrógeno en el suelo, empleando la espectroscopia NIR y modelos con diferentes enfoques de aprendizaje, incluyendo vectores de soporte y arboles de decisión. Donde también se investiga el uso del aprendizaje profundo, con la implementación de redes generativas adversarias para aumentar la cantidad y calidad de los datos disponibles para entrenamiento de los modelos, lo que potencialmente mejora la precisión de las predicciones.

Los resultados obtenidos a través de los experimentos realizados con estas técnicas y enfoques sustentan a este trabajo académico como una contribución al conocimiento en la intersección de tecnología y agricultura, al demostrar cómo la integración de técnicas avanzadas de análisis puede revolucionar la manera en que se gestionan los nutrientes del suelo para una producción agrícola más eficiente y sostenible.

Palabras clave: Elementos del suelo, Teledetección, Análisis espectral, Algoritmos computacionales, Modelos de predicción.

¹ Tesis de Doctorado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo

Autor: Alejandro Eric Reyes Rivera

Director de Tesis: Dr. Gilberto de Jesús López Canteñs

GENERAL ABSTRACT

Soil nutritional analysis with Machine Learning and Vis-NIR Spectroscopy²

The present thesis is focused on the application of Machine learning techniques (ML) and spectral analysis using Vis-NIR spectroscopy to enhance soil nutrient analysis within the context of precision agriculture. The primary objective is to optimize accuracy and efficiency in predicting soil nutrients, particularly nitrogen. Two essential approaches are addressed in the thesis: variable selection and ML algorithm calibration. On one hand, the effectiveness of three methods for selecting key wavelength intervals in predicting total soil nitrogen and pH is explored. This is achieved by employing Vis-NIR spectral data and search and optimization methods, including bio-inspired algorithms.

On the other hand, a more detailed examination is conducted regarding calibration methods for soil nitrogen prediction. This involves employing NIR spectroscopy and models with different learning approaches, including support vector machines and decision trees. Furthermore, the use of deep learning is also investigated, involving the implementation of Generative Adversarial Networks (GANs) to enhance both the quantity and quality of available data for model training. This has the potential to enhance prediction accuracy.

The results obtained through experiments conducted using these techniques and approaches substantiate this academic work as a contribution to knowledge at the intersection of technology and agriculture. It demonstrates how the integration of advanced analytical techniques has the capacity to revolutionize the management of soil nutrients for a more efficient and sustainable agricultural production.

Keywords: Soil elements, Remote sensing, Spectral analysis, Computational algorithms, Prediction models.

²Thesis, Universidad Autónoma Chapingo
Author: Alejandro Eric Reyes Rivera
Advisor: Dr. Gilberto de Jesús López Canteñs

CAPITULO 1: INTRODUCCIÓN GENERAL

El análisis nutricional de suelos es fundamental para comprender la disponibilidad de macronutrientes esenciales como el nitrógeno, el fósforo y el potasio (NPK), los cuales son vitales para el crecimiento de las plantas y también constituyen los principales componentes de los fertilizantes. La reducción de los niveles óptimos de contenido de NPK y otros nutrientes provoca una disminución de productividad del suelo (G. Xu, *et al*, 2013), mientras que su aplicación en exceso puede tener un impacto ambiental negativo. Por lo tanto, la determinación del contenido de NPK y su administración por sitio-específico en el suelo desempeñan un papel fundamental en la agricultura moderna (Patel, *et al*, 2020). La agricultura de precisión (AP) se ha beneficiado de la integración de tecnologías disruptivas, como las basadas en inteligencia artificial (IA) y la espectroscopia (Kose *et al.*, 2022). Estas herramientas han permitido mejorar la eficiencia en la gestión de recursos agrícolas, optimizando insumos y aumentando la productividad de los cultivos. En AP, los sistemas de detección remota o teledetección (Mirzaee *et al.*, 2016) tienen como finalidad obtener información sobre la superficie terrestre sin contacto, a través de sensores montados en vehículos aéreos, como satélites, drones o aviones. En el análisis de fertilidad del suelo, la teledetección comenzó a implementarse en la década de 1960 (Lei, *et al*, 2011), Sin embargo, su impacto significativo en la obtención de información sobre la composición química del suelo y la toma de decisiones agrícolas más informadas se ha evidenciado en los últimos años (Ahmadi *et al.*, 2021; Barra *et al.*, 2021; Laskar & Mukherjee, 2016; Soriano-Disla *et al.*, 2014). Actualmente, los sensores utilizados en teledetección agrícola, como los espectrorradiómetros, sensores de fluorescencia y sensores de infrarrojo térmico, tienen fundamento en la espectroscopía. Esta técnica ofrece una forma efectiva de detectar y cuantificar nutrientes en el suelo a través de sus características de absorción y reflectancia únicos en diversas longitudes de onda (Isaac & Na, 2017). La espectroscopía, complementada con técnicas de análisis de IA ha revolucionado la capacidad de obtener detalles sobre la composición química de suelos y plantas, mejorado la interpretación de datos, ya que permiten relacionar

la reflectancia (o absorbancia) del suelo con elementos clave para la agricultura a través del análisis espectral (Inoue, 2020; Laskar & Mukherjee, 2016). Estudios previos han validado la utilidad de los rangos espectrales visible (Vis), infrarrojo cercano (NIR) y medio (MIR) para evaluar nutrientes en suelos agrícolas (Nocita *et al.*, 2014; Shao & He, 2011; Soriano-Disla *et al.*, 2014; Vohland *et al.*, 2014; H. Yang *et al.*, 2012). En estos estudios, subcampos de la IA como el aprendizaje automático (ML por Machine Learning), desempeñan un papel central en el desarrollo de modelos de reflectancia de nutrientes en el suelo. Estas técnicas permiten procesar grandes volúmenes de datos espectrales y capturar relaciones complejas, adaptándose a diversas condiciones y contextos (Ahmadi *et al.*, 2021).

El desarrollo de modelos espectrales es desafiante debido a la complejidad de los espectros, la presencia de ruido espectral, información redundante e irrelevantes (Petroopoulos *et al.*, 2012), y las alteraciones causadas por factores como la textura y humedad de la muestra. Para abordar estos desafíos y avanzar en este análisis, la comunidad científica se centra en aspectos clave en el proceso de desarrollo de modelos espectrales. Estos incluyen la búsqueda de pretratamientos efectivos para los espectros de suelos (Buddenbaum & Steffens, 2012; P. C. Nie *et al.*, 2017; Yubing Wang *et al.*, 2021); la determinación de longitudes de onda características de cada elemento (Kawamura *et al.*, 2021; Min & Lee, 2005; Yuan *et al.*, 2020) y la evaluación de algoritmos de regresión apropiados para relacionar las propiedades con la reflectancia o absorbancia del suelo (Ding *et al.*, 2018; Morellos *et al.*, 2016; R. Vestergaard *et al.*, 2021; M. Yang *et al.*, 2019). Sin embargo, la efectividad de los modelos de ML depende de la cantidad y calidad de los datos espectrales usados para el entrenamiento (Ben-Dor *et al.*, 2009; Qi *et al.*, 2018). Sin embargo, la obtención de datos reales es costosa y a menudo carece de variedad. Debido a esto, la generación de espectros sintéticos se ha convertido en una estrategia prometedora para aumentar el conjunto de datos disponible para el entrenamiento de modelos predictivos (Jiang *et al.*, 2023). Esto implica el uso de técnicas de simulación con

modelos de aprendizaje profundo que permiten crear datos virtuales con características similares a los datos reales (Goodfellow *et al.*, 2020).

En este trabajo, se estudian las bases y principios del aprendizaje automático y la espectroscopia como herramientas fundamentales en la agricultura moderna para mejorar la gestión de macronutrientes y promover la sostenibilidad agrícola. A continuación, se describen los objetivos de investigación del proyecto:

Objetivo Principal: Estudiar, desarrollar e implementar metodologías innovadoras que integren modelos de aprendizaje automático y técnicas de espectroscopía Vis-NIR, para determinar con precisión los niveles de nutrientes en suelos en el contexto de la agricultura de precisión.

Particularmente:

- Identificar y evaluar métodos de aprendizaje automático enfocados al pretratamiento y selección de variables en espectroscopia de suelos, con la finalidad de analizar longitudes de onda relevantes en la determinación de nitrógeno y pH de suelos, utilizando una biblioteca espectral Vis-NIR.
- Comparar la capacidad predictiva de modelos más utilizados en ciencias de suelos, desarrollados a partir de algoritmos de aprendizaje automático y una biblioteca espectral NIR de muestras del campo experimental “La Xerona” del Departamento de Ingeniería Mecánica Agrícola, con el propósito de determinar los enfoques de regresión más adecuados para nitrógeno.
- Evaluar la utilidad de modelos generativos de aprendizaje profundo para mejorar la calidad y variabilidad de la biblioteca espectral desarrollada para la detección de nitrógeno de suelos.

Los siguientes 3 capítulos comprenden: en el Capítulo 2, el marco teórico de los temas Suelo, Espectroscopia y Aprendizaje Automático; en el Capítulo 3, el artículo: “Evaluación de tres métodos de selección de variables para predicción de nitrógeno total y pH del suelo usando datos espectrales Vis-NIR”; finalmente en el Capítulo 4, el artículo: “Evaluación de algoritmos de

calibración para predicción de nitrógeno del suelo mediante espectroscopía NIR y aumento de datos con redes generativas adversarias”.

CAPITULO 2: MARCO TEÓRICO

2.1 MACRONUTRIENTES DEL SUELO

El suelo es un material complejo con una composición física y química altamente variable, influenciada por factores como el clima, la vegetación, la fauna, la topografía y la roca madre (Wilson, 2019). Se puede caracterizar como un sistema trifásico que comprende una fase sólida, líquida y gaseosa (Figura 2.1). La fase sólida contiene componentes orgánicos e inorgánicos, incluyendo minerales, materia orgánica y sales (Ehrenfeld *et al.*, 2005). Está constituida a partir de tres tamaños de partícula principales: arena (2-0,2 mm), limo (0,2-0,002 mm) y arcilla (<0,002 mm), que juntas definen dos propiedades principales del suelo: su textura, como la distribución del tamaño de partícula, y la estructura, que está relacionada con las fuerzas de adhesión entre partículas. Estas propiedades juegan un papel importante en algunas propiedades del suelo, como porosidad, drenaje, fertilidad, retención de humedad y erosión (Bronick & Lal, 2021).

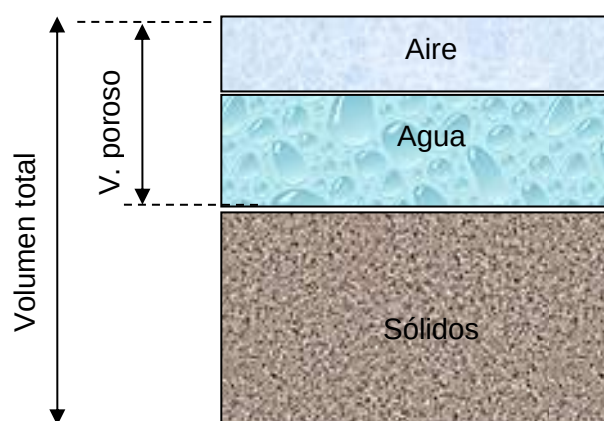


Figura 2.1. El suelo como sistema trifásico.

Por otro lado, la fase gaseosa del suelo refleja la composición atmosférica local, con fluctuaciones en la concentración de oxígeno y dióxido de carbono debido a la actividad biológica en la zona de raíces. Por último, la fase líquida

principalmente agua, interactúa con el suelo de diferentes maneras: llenando por completo los poros en suelos saturados, ocupando una parte de los poros en suelos húmedos o siendo adsorbida en áreas superficiales en suelos secos. La suma de la fase líquida y gaseosa constituye el 'volumen poroso', que influye en la retención de agua y el intercambio de gases esenciales para las plantas y la biología del suelo (Ehrenfeld *et al.*, 2005).

El suelo además consta de propiedades químicas (nutrientes) y biológicas que brindan la capacidad de sustentar la producción de plantas. Los nutrientes del suelo desempeñan un papel vital en el desarrollo de cultivos agrícolas, y se clasifican en dos categorías principales: macro y micronutrientes. Los micronutrientes, que son requeridos en pequeñas cantidades, incluyen elementos como el hierro (Fe), el manganeso (Mn), el cobre (Cu), el zinc (Zn), el boro (B), el molibdeno (Mo) y el cloro (Cl). Son esenciales para diversas funciones vitales en las plantas, como la activación de enzimas, la síntesis de clorofila y la regulación de procesos metabólicos. La carencia de micronutrientes puede provocar deformidades en las hojas, baja calidad de granos, menor resistencia a enfermedades y plagas, así como disminución en el rendimiento de los cultivos (Dhaliwal *et al.*, 2019).

Los macronutrientes por otro lado incluyen los elementos nitrógeno (N), fósforo (P) y potasio (K), que se consideran como primarios y, calcio (Ca), magnesio (Mg) y azufre (S) que son secundarios. Todos los macronutrientes son necesarios en grandes cantidades para el crecimiento y desempeñan funciones clave en el desarrollo de las plantas, como la formación de proteínas, el fortalecimiento de las estructuras celulares, la producción de energía y la regulación del equilibrio de agua. La deficiencia de macronutrientes puede manifestarse en síntomas visuales, como hojas amarillentas o pálidas, retraso en el crecimiento y una menor producción de frutos y granos (Kim *et al.*, 2009).

Los elementos químicos NPK son los más importantes de la materia orgánica vegetal presente en el suelo y también constituyen los principales componentes de los fertilizantes. La reducción de los niveles de NPK y otros nutrientes, puede resultar en una disminución de fertilidad, porosidad, penetrabilidad y en

consecuencia, productividad del suelo (G. Xu, *et al*, 2013). Por lo que la determinación cuantitativa de estos elementos y su variabilidad intraparcelsaria es clave para el control de procesos de metabolismo y la salud de suelos y cultivos). Además, este manejo debe ser sostenible ambientalmente. Una aplicación excesiva de estos elementos en el suelo perjudica la calidad de los cultivos y conduce a un desequilibrio ecológico importante, contaminando los cuerpos de agua y las aguas subterráneas por percolación profunda (Weiss *et al*, 2020).

2.2 DETECCIÓN DE PROPIEDADES NUTRICIONALES EN SUELOS

En Agricultura de Precisión (AP), se emplean técnicas de teledetección que permiten evaluar rápidamente las condiciones de los cultivos y algunas características del suelo, facilitando así la toma de decisiones en la gestión agrícola. La teledetección es un conjunto de métodos que cuantifican propiedades de objetos a distancia, utilizando sistemas basados en sensores ópticos, imágenes satelitales y muestreo intensivo del suelo. Según Gonzalez & Vargas, (2014), un proceso de teledetección involucra varios elementos clave, desde la fuente de energía o radiación, generalmente proporcionada por el Sol, los sensores, su interacción con los objetos de interés, transmisión, recepción, hasta la interpretación y análisis de los datos recopilados. En estos sistemas, los sensores ópticos desempeñan un papel fundamental al captar la energía reflejada en diversas longitudes de onda del espectro electromagnético, como el visible (Vis), el infrarrojo (IR) y las microondas (MW). Estos sensores se montan en diferentes plataformas, como satélites, aeronaves, drones (UAV/UAS) o tractores (Thenkabail *et al.*, 2018).

En la práctica, se han propuesto una serie de índices espectrales como el NDVI, EVI, SAVI que utilizan la información espectral de algunas longitudes de onda para evaluar el estado del cultivo y del suelo. Sin embargo, estos índices presentan limitaciones en términos de sensibilidad, resolución espacial y capacidad de interpretación, pues aspectos clave como la nutrición del suelo no se puede obtener directamente con estos enfoques (Inoue, 2020). Para tal

propósito, tradicionalmente se dispone de métodos estandarizados de laboratorio como el método de Kjeldahl (Sáez-Plaza *et al.*, 2019) para N, los métodos de extracción de Bray y Kurtz (Bray & Kurtz, 1945) y de Olsen (Horta & Torrent, 2007) para el caso del P, así como métodos de extracción con acetato de amonio (NH₄OAc) o ácido clorhídrico (HCl) para el contenido de K (Barefoot *et al.*, 2003). Sin embargo, estos métodos tienen muchas limitantes respecto al costo, el tiempo de análisis y la gran variabilidad de los suelos, por lo que resultan viables para su implementación en AP.

En ese contexto, con el avance tecnológico han surgido metodologías de análisis más sofisticadas que utilizan una exploración espectral más profunda que los índices espectrales, donde la integración de Inteligencia Artificial y sus distintas ramas como el Aprendizaje automático han tomado un papel determinante. Lo cual permite una detección más precisa, un monitoreo más detallado y una toma de decisiones más informada. Estas técnicas, bajo las condiciones y metodología adecuadas, han logrado obtener resultados positivos en la descripción cuantitativa y cualitativa de diversas propiedades del suelo (Barra *et al.*, 2021; J. Liu *et al.*, 2020; Nawar *et al.*, 2019; Soriano-Disla *et al.*, 2014). Sus aplicaciones y utilidad se extienden con el uso de imágenes multiespectrales e hiperespectrales (H. Li *et al.*, 2019; J. Ma *et al.*, 2022; Patel *et al.*, 2020; Qi *et al.*, 2018). Así como con el desarrollo de prototipos de detección y monitoreo del suelo en el sitio (Ahmadi *et al.*, 2021; Isaak *et al.*, 2019; Ji, Adamchuk, *et al.*, 2016; Y. Li *et al.*, 2019; Zhou *et al.*, 2019).

2.3 ESPECTROSCOPIA

La Espectroscopía se ha destacado como una técnica fundamental en la Agricultura de Precisión y en el análisis de suelos. Debido a sus ventajas y características que fomentan una toma de decisiones más fundamentada y una gestión más efectiva de los recursos, lo que, en última instancia, promueve una agricultura más sostenible y económicamente viable. Por lo tanto, es esencial comprender algunos conceptos básicos relacionados con esta técnica.

2.3.1 Generalidades

Es una técnica empleada para identificar tanto cualitativa como cuantitativamente los componentes de diversas sustancias o mezclas, tanto en estados estáticos como dinámicos (González y Montaña, 2015). Una herramienta ampliamente utilizada en la investigación, el análisis, el control y el diagnóstico en diversas disciplinas, incluyendo la física, la química y las ciencias biológicas. Su aplicabilidad se ha traducido en avances tecnológicos dirigidos al análisis y monitoreo en diversas industrias, incluida la agricultura.

Las ventajas técnicas de los métodos espectroscópicos son notables, destacándose su naturaleza no destructiva y la capacidad de estudiar objetos de interés sin contacto físico, lo que los hace ideales para aplicaciones de teledetección. Además, requieren una preparación mínima de las muestras, evitando la necesidad de tratamientos químicos y reactivos, lo que resulta en la reducción de costos y un beneficio ecológico significativo (Linker *et al.*, 2005). Sin embargo, el uso de la espectroscopía también conlleva desafíos, como el ruido y otras perturbaciones en la respuesta espectral causadas por la dispersión de radiación, lo que complica el proceso de análisis. Los equipos espectroscópicos modernos generan espectros con una gran cantidad de variables e información redundante, lo que requiere técnicas específicas para depurar los datos. Para abordar estos desafíos, en las últimas décadas, el crecimiento de la inteligencia artificial, en particular el aprendizaje automático, ha impulsado avances significativos en la aplicación de la espectroscopía, lo que ha llevado a una mejora notable en la eficiencia y precisión de los procesos de análisis (Barra *et al.*, 2021; Weiss *et al.*, 2020). Estos avances incluyen la utilización de pretratamientos espectrales para resaltar características de reflectancia y eliminar perturbaciones en las mediciones. Además, se han implementado técnicas avanzadas de optimización y análisis multivariante para seleccionar la información más relevante y establecer correlaciones precisas entre la reflectancia y los elementos de interés, como se demuestra en el trabajo de Lin *et al.*, (2017). Estos avances representan un hito importante en la evolución de la

espectroscopía, proporcionando herramientas más potentes y efectivas para una variedad de aplicaciones.

2.3.2 Principios

La palabra espectro da lugar a la raíz etimológica de esta técnica, que se complementa con el sufijo “scopia” del griego “skopia” que se traduce como “acción de ver, observar o analizar”. Mas ampliamente, se considera a la espectroscopia como el estudio de la interacción (emitancia, absorbancia y transmitancia) entre radiación electromagnética y materia (Donald *et al.*, 2017). Consecuentemente, se puede distinguir entre espectroscopia de absorción y espectroscopia de emisión. Aunque estas pueden diferir significativamente en técnicas de medición, aplicaciones e interpretación de resultados, en ambos casos se mide intensidad de energía en forma de radiación electromagnética (Vanderlinde y Neuenschwander, 1994). A continuación, se profundiza en los conceptos básicos más importantes.

Radiación electromagnética

La radiación electromagnética es una forma de energía que se produce naturalmente debido a transiciones energéticas de los electrones dentro de los átomos. En el espacio, esta energía se propaga en forma de onda y está compuesta por dos campos; uno eléctrico y otro magnético, los cuales oscilan sinusoidalmente, dando paso al concepto de onda o campo electromagnético. Estos campos son perpendiculares entre sí y a la dirección de propagación de la propia onda en cualquier instante del tiempo (Manolakis *et al.*, 2016).

En el vacío, la radiación electromagnética viaja a la velocidad de la luz (C), y puede caracterizarse por su longitud de onda λ ($mm, \mu m, nm$), por su número de onda $\tilde{\nu}$ (cm^{-1}) o por su frecuencia ν (GHz). La relación intrínseca que existe entre estas características espectrales es la siguiente:

$$\lambda = \frac{C}{\nu} = \frac{1}{\tilde{\nu}} \quad (2.1)$$

La radiación electromagnética incluye, además de lo que comúnmente llamamos "luz", radiación de longitudes de onda más largas y cortas, las cuales en conjunto

conforman el espectro electromagnético. La radiación de determinada región espectral y su interacción con la materia originan las diferentes técnicas espectroscópicas utilizadas en la actualidad (Hollas, 2011).

2.3.2.2 Interacción entre radiación y materia

Existen tres fenómenos principales de interacción entre materia y radiación conocidos como; absorción inducida, emisión espontánea y emisión estimulada (Buchmiller & Tye, 1980). En la Figura 2.2, se ilustran dos estados estacionarios de las moléculas de un sistema (materia) sometido a radiación de frecuencia $\tilde{\nu}$. Se tiene el estado natural E_m y el estado inducido E_n , la diferencia entre estos estados corresponde a un cambio de energía $hc\tilde{\nu}$ en el sistema, es decir:

$$\Delta E = E_n - E_m = hc\tilde{\nu} \quad (2.2)$$

Este cambio de energía da paso a los siguientes procesos de interacción:

1. **Absorción inducida.** Ocurre cuando una molécula o átomo M es estimulada con radiación, pasando de un estado natural a un estado artificial M^* , es decir: $M + hc\tilde{\nu} \rightarrow M^*$
2. **Emisión espontánea.** En este proceso, una molécula parte inicialmente de un estado n y espontáneamente emite o pierde energía (radiación), esto es: $M^* \rightarrow M + hc\tilde{\nu}$
3. **Emisión inducida.** Es una forma de emisión de energía en la cual, primero se requiere inducir M a una condición de alteración M^* , para después provocar emisión de energía y pasar de estado n al m , dicho de otra forma: $M^* + hc\tilde{\nu} \rightarrow M + 2hc\tilde{\nu}$

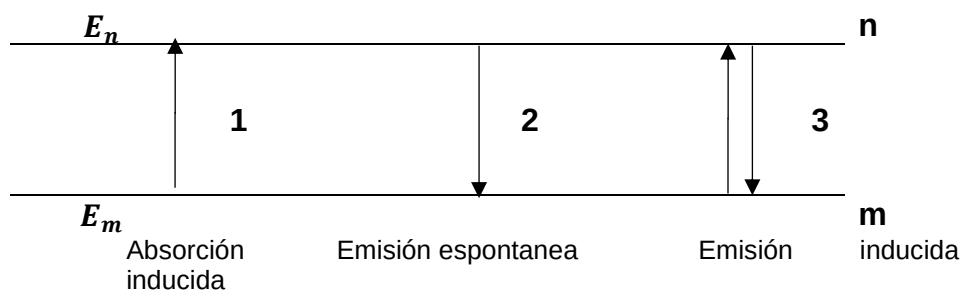


Figura 2.2 Procesos de absorción y emisión en la interacción materia-radiación

De acuerdo con Hollas, (2011) una molécula puede experimentar diferentes estados de excitación, dependiendo del tipo de radiación inducida sobre ella, a estas transmisiones de energía se les conoce como estados vibracionales (EV). El Cuadro 2.1 resume los principales estados vibratorios de acuerdo con la región electromagnética implicada.

Los hay del tipo rotacional, que responden a radiación con longitudes de onda de baja frecuencia, estos son típicamente utilizados en los procesos de cocción con microondas, o en la radio. Los EV de tipo vibratorios por otro lado, se presentan al interactuar con radiación infrarroja y visible. Hay dos clases de vibraciones básicas o fundamentales; de estiramiento ("stretching") y de deformación o flexión (bending). De aquí se desprenden técnicas como la espectroscopia visible (VIS) e infrarroja (IR).

Cuadro 1.1. Estados vibratorios en función de la región espectral.

Región	Longitud de onda	Estado vibratorio
Rayos X	<280 nm	De ruptura de enlaces
Ultravioleta	280-400 nm	Electrónico
Visible	400-800 nm	Electrónico
infrarrojo	770 nm - 0.5 mm	Vibracional
Microondas	10 mm - 1m	Rotacional
Radio frecuencias	1 - 100 m	Espín: nuclear/electrónico

En espectroscopía VIS, además de las vibraciones fundamentales, se presentan transiciones electrónicas, también llamados saltos cuánticos a nivel molecular y atómico. Por otro lado, la espectroscopía IR se puede segmentar principalmente en espectroscopía de infrarrojo medio (MIR) y de infrarrojo cercano (NIR). En el MIR y NIR aparecen superposiciones y combinaciones de las vibraciones básicas.

Asimismo, otras técnicas como las de Resonancia Magnética Nuclear (RMN) y Resonancia de Espín (ESR) involucran efectos de espín (Uhlenbeck & Goudsmit, 1976), de tipo nuclear y electrónico, respectivamente.

Por otro lado, existen técnicas espectroscópicas que se definen más de acuerdo al proceso de transmisión de energía que a su fuente de radiación, un ejemplo claro es la Espectroscopia de Raman (Lyon *et al.*, 1998), en donde la radiación experimenta cambios en su longitud de onda una vez que interactúa con las moléculas (efecto Raman) (Singh, 2002), cuya fuente suele estar en la región visible o casi ultravioleta del espectro electromagnético.

En la Figura 2.3 se representa una representación de la extensión del espectro electromagnético, desde las ondas de radio de baja frecuencia hasta la radiación de rayos gama de alta frecuencia.

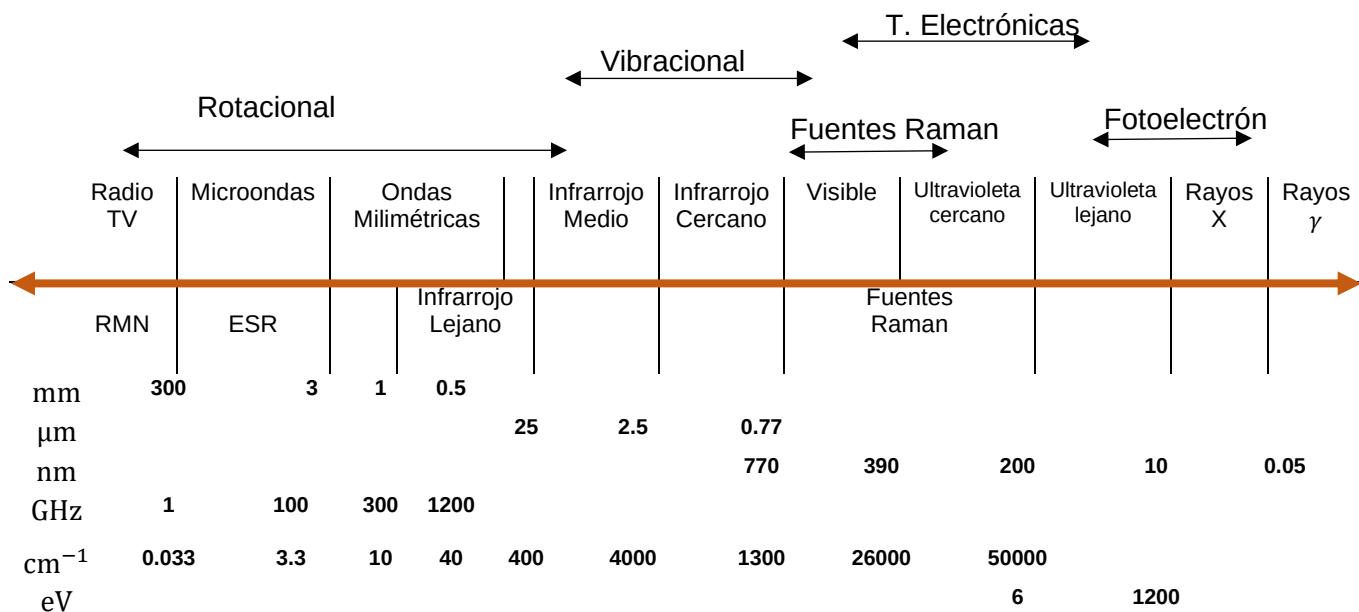


Figura 2.3 Espectro electromagnético y regiones importantes.

2.3.2.3 La ley de Beer-Lambert en medios difusos

La ley de Beer es clave en el análisis espectroscópico NIR cuantitativo. Se basa en la hipótesis de que existe una relación entre la respuesta espectrométrica (reflectancia, absorbancia o transmitancia) y la concentración de un elemento o analito en una muestra (a). Se supone que el cociente entre la intensidad transmitida $I_T(\lambda)$ en un medio y la intensidad de la fuente de radiación incidente $I_0(\lambda)$ es equivalente a:

$$T(\lambda) = \frac{I_T(\lambda)}{I_0(\lambda)} = 10^{-\varepsilon(\lambda) \cdot a \cdot l} \quad (2.3)$$

Donde $T(\lambda)$ es la transmitancia en determinada longitud de onda λ , $\varepsilon(\lambda)$ es la absortividad o coeficiente de extinción molar (capacidad que tienen las especies químicas para absorber radiación en una determinada longitud de onda, $L \cdot mol^{-1} \cdot cm^{-1}$), l es la longitud de la trayectoria del haz de luz.

En absorbancia, se aplica el logaritmo para linealizar la relación entre la respuesta espectral y la concentración:

$$A(\lambda) = -\log\left(\frac{I_T(\lambda)}{I_0(\lambda)}\right) = \varepsilon(\lambda) \cdot a \cdot l \quad (2.4)$$

Aquí, el término $\varepsilon(\lambda) \cdot c$ caracteriza la capacidad de absorción de la muestra analizada y puede ser reemplazada por el coeficiente de absorción, es decir: $\mu(\lambda) = \varepsilon(\lambda) \cdot a$. Este coeficiente representa la probabilidad por unidad de longitud, que tiene un fotón de determinada λ , de ser absorbido por el material con el que interacciona (Gobrecht *et al.*, 2014).

Debido a que el suelo es altamente absorbente y dispersante por su composición física (textura), se considera un medio difuso donde no se pueden satisfacer las condiciones teóricas ideales que exige la ley de Beer-Lambert. En un medio difuso se produce además de la absorción, el fenómeno de dispersión. La dispersión se refiere al cambio de trayectoria del fotón al chocar con una partícula o cuando cambia el índice de refracción (Ciani *et al.*, 2005). Las no linealidades en las relaciones de reflectancia (o absorbancia), con los elementos del suelo originadas por la dispersión, son un desafío para la calibración multivariante. La ley de Beer-Lambert se aplica en medios de difusos reemplazando $I_T(\lambda)$ por la intensidad de la radiación remitida $I_R(\lambda)$ y el coeficiente de absorción $\mu_a(\lambda)$ por el de dispersión $\mu_s(\lambda)$. La expresión analítica del coeficiente $\mu_s(\lambda)$ es compleja, porque los cambios de dirección de los fotones dependen no solo del tamaño y la forma de las partículas, sino también de la longitud de onda, la dirección de la luz incidente y las variaciones de índices refractivos en cada medio analizado (Gobrecht *et al.*, 2014).

En la Figura 2.4A se muestra el camino óptico ideal de la radiación en una medición de transmitancia en muestras homogéneas. El aumento en la longitud

del camino recorrido por los fotones en un medio disperso refleja dos efectos importantes que pueden alterar la respuesta espectral:

- Efecto multiplicativo en absorbancia (Figura 2.4B). Cuando la dispersión provoca un aumento de la longitud del camino óptico de un fotón, este tiene más probabilidades de ser absorbidos por el medio, lo cual induce altas variaciones de absorbancia en muestras de la misma naturaleza.
- Efecto aditivo en absorbancia (Fig. 4.1C). Es un efecto contrario al multiplicativo. La dispersión además de aumentar el camino óptico de los fotones puede variar su dirección de salida, originando pérdidas en la respuesta espectroscópica, debido a que los instrumentos ópticos están contruidos de tal manera que solo se detecta una fracción de radiación transmitía por la muestra (Gobrecht *et al.*, 2014).

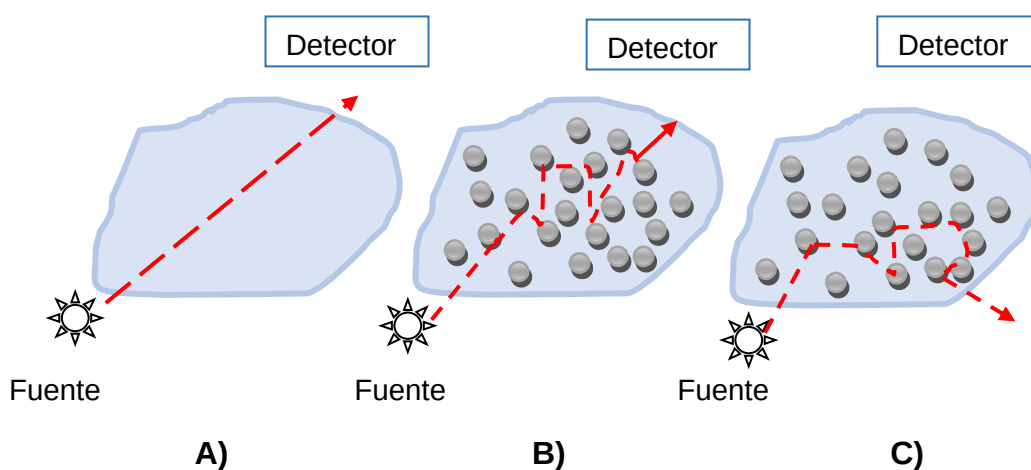


Figura 2.4. Efectos de dispersión en medios no difusos (A) y difusos (B y C).

Finalmente, la presencia de efectos multiplicativos y aditivos dominantes en los datos espectrales implica una variabilidad entre muestras, que da como resultado desviaciones de la línea de base del espectro ideal (Gobrecht *et al.*, 2014). Además, podrían invalidar la suposición subyacente de los métodos de calibración lineales multivariantes de uso común. Los cuales postulan una relación lineal entre las mediciones espectrales y los componentes químicos en una muestra. Por lo tanto, el rendimiento predictivo de los modelos multivariantes lineales podría deteriorarse significativamente (Jin *et al.*, 2012).

2.3.3 Espectroscopía para suelos

En el sector agrícola, las características fisicoquímicas del suelo determinan su utilidad práctica, por lo que la espectroscopia para detección remota en suelos es un pilar muy importante en AP (Inoue, 2020). En este apartado se describen las técnicas comúnmente utilizadas en análisis espectrométricos de suelos, los equipos frecuentemente utilizados para la obtención de espectros, así como algunos aspectos clave para la interpretación de información contenida en los espectros del suelo.

2.3.3.1 Espectroscopia Infrarroja con Transformada de Fourier FT-IR

La base de los sistemas espectroscópicos se fundamenta en la diferencia entre dos magnitudes de radiación; la energía irradiada a la muestra y la detectada después de su interacción con la materia. La espectroscopía que opera en el rango espectral infrarrojo IR es la más utilizada en análisis de suelos, debido a que la mayoría de las moléculas presentes reaccionan a radiación de frecuencias de 600-25000 nm (Papousek y Aliev, 1982).

En un inicio, las muestras se irradiaban con diferentes λ únicas a través de un proceso repetitivo, tardado y de menor precisión, para lo cual se utilizaban instrumentos de medición llamados espectrómetros infrarrojos dispersivos.

La espectroscopia con transformada de Fourier FT-IR (Bartholdi y Ernst, 1973), permitió recoger los datos espectrales de todas las λ en una sola etapa. En la figura 2.5 se muestra un diagrama de funcionamiento de los equipos espectroscópicos FT-IR y a continuación se describe el proceso de acuerdo con (Donald et al., 2017):

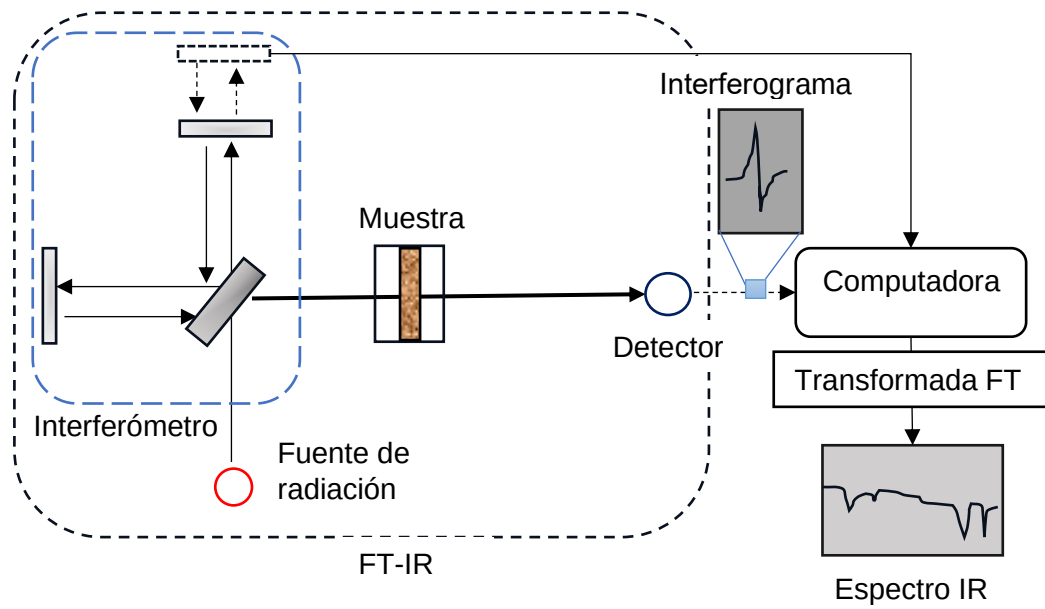


Figura 2.5. Diagrama de un Espectroscopio FT-IR

1. **Fuente de radiación:** Fuente que emite luz en un amplio rango de frecuencias infrarrojas. Puede ser una fuente de globar (filamento de tungsteno) o una fuente de lámpara de tungsteno-halógeno.
2. **Interferómetro:** Instrumento óptico que procesa la radiación antes de pasar por la muestra de la siguiente manera: La radiación entrante pasa a través de un **divisor de haz**, un espejo colocado en un ángulo de 45° , que la separa en dos haces perpendiculares, uno sin desviar, el otro orientado a 90° . Este haz, va a un **espejo fijo** mientras que el haz no desviado va a un **espejo móvil** que hace que varíe la longitud de su trayectoria, ambos regresan al divisor de haz. Cuando los dos haces se encuentran en el divisor de haz, se recombinan, pero las diferencias de longitud de trayectoria (diferentes λ) provocan interferencias tanto constructivas como destructivas. El haz combinado que contiene estos patrones de interferencia se denomina interferograma, para obtenerlo se usa el principio de interferencia de ondas (Atasoy *et al.*, 2011), que contiene toda la energía de la fuente y una amplia gama de λ en el dominio del tiempo.
3. **Muestra:** El interferograma se orienta hacia la muestra mediante el divisor de haz. La muestra absorbe simultáneamente todas las λ en su espectro

infrarrojo. La señal de interferograma modificada es enviada al detector con información sobre la cantidad de energía que se absorbió en cada λ .

4. **Detector:** El detector mide la intensidad de luz transmitida o absorbida en función de la frecuencia infrarroja. El detector puede ser un detector de interferencia o de estado sólido sensibles al infrarrojo, como detectores de mercurio-cadmio-telurio (MCT) o detectores de bolómetro.
5. **Registro de datos:** Una computadora compara el interferograma modificado con un haz de referencia. Estos datos se convierten en un espectro infrarrojo aplicando una transformada de Fourier, que muestra las intensidades de luz en función de la frecuencia o longitud de onda.

Procesamiento de Fourier: Al contener la información en el dominio del tiempo, el interferograma no puede ser leído de manera directa. Se debe implementar un proceso matemático llamado transformada de Fourier. La transformada de Fourier lo descompone en sus componentes de frecuencia y proporciona el espectro infrarrojo final. A través de las series de Fourier (2.5), una función de dominio temporal o espacial $f(x)$ se puede representar en un dominio frecuencial, por medio de un conjunto de sumatorias de senos y cosenos que describen la señal (Brunton y Kutz, 2019).

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) + b_n \sin(nx) \quad (2.5)$$

En la ecuación 2.5 se tienen diferentes componentes frecuenciales (λ) determinadas por n , la amplitud de los senos y cosenos en la función frecuencial está dada por los coeficientes a y b respectivamente, los cuales deben ser determinados para cada caso. Debido a que es necesario trasladarse a un dominio frecuencial, la ecuación 2.5 debe ser reformulada en su representación compleja, utilizando las propiedades descritas en 2.6 y 2.7:

$$e^{ix} = \cos(x) + i\sin(x) \quad (2.6)$$

$$e^{-ix} = \cos(x) - i\sin(x) \quad (2.7)$$

De donde se obtiene la versión trasladada al plano complejo de la transformada de Fourier (2.8), con la cual trabaja la espectroscopia FT.

$$f(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\lambda x} dx \quad (2.8)$$

Donde $f(x)$ = Señal en un dominio temporal-espacial x , λ = Número de onda y $e^{-i\lambda x} = \cos(\lambda x) + i\sin(\lambda x)$.

2.3.3.2 Métodos espectroscópicos para suelos

Existen diversos métodos de obtención de espectros con espectroscopia FT en sus diferentes dominios de operación (NIR, MIR, Vis-NIR). Algunos de los métodos más utilizados para obtener espectros de suelo son: La Reflectancia Especular, la Reflectancia Total Atenuada y Reflectancia Difusa (Figura 2.6). A continuación, se describen estas técnicas en sus principios básicos.

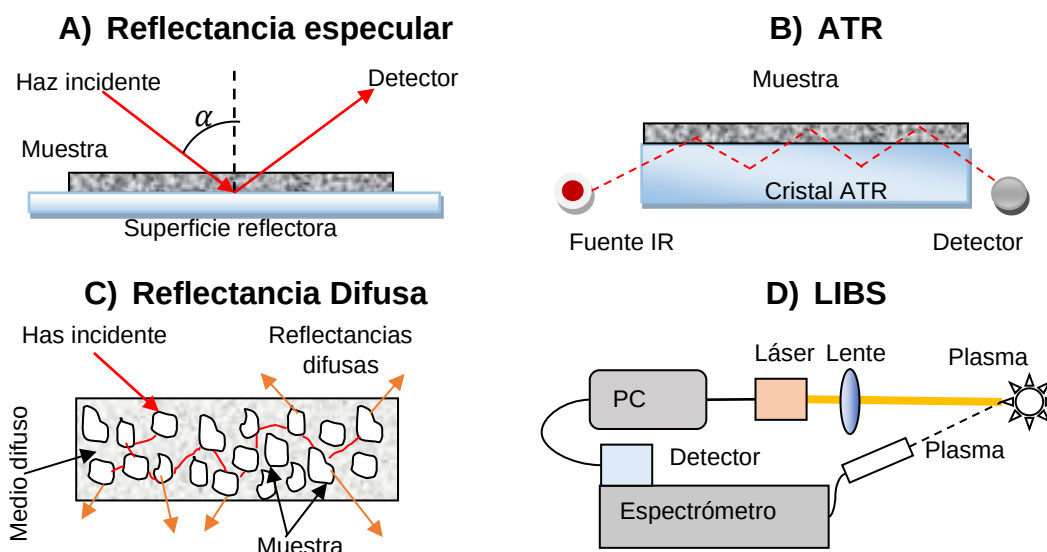


Figura 2.6. Diagramas de los métodos de medición de espectros IR de suelos

- **Espectroscopia FT con Reflectancia Especular:** También conocida como espectroscopia de absorción y reflexión (RAS), está basada en el efecto Rayleigh (Young, 1981). Se utiliza una superficie reflectiva como contenedor, sobre la cual se coloca una fina capa de muestra (Figura 2.6A). Entonces, un haz de radiación es dirigido desde a fuente con un ángulo de ataque determinado (α). La radiación pasa a través de la muestra hacia la superficie, generando la absorción, una vez en contacto con la superficie, la radiación es reflejada hacia el detector. Una de las

variables más importantes de este método es el ángulo α , pues dependiendo de su valor se determina la sensibilidad de las mediciones, al regular la cantidad de muestra que atraviesa el haz de radiación. Este ángulo es frecuentemente controlado por un modulador digital que lo regula para obtener mejores resultados en el análisis espectral.

- **Espectroscopia FT con Reflectancia Total Atenuada:** La técnica estándar para la medición de espectros FT es la espectroscopia de reflectancia total atenuada (ATR). En este método se emplea además de los componentes ya mencionados, un cristal que puede ser de diamante, *ZnSe* o Germanio (*Ge*). Este cristal debe tener un índice refractivo superior a la muestra estudiada. Durante la medición, la radiación es dirigida hacia el cristal que presiona la muestra y entonces es reflejada internamente (Figura 2.6B), entonces la radiación es parcialmente absorbida por la muestra según sus propiedades espectrales. La energía sobrante es reflejada hacia el detector, la diferencia de radiación emitida y reflejada contiene la información de absorción de la muestra.
- **Espectroscopia FT con Reflectancia Difusa:** En espectroscopia difusa se emplea un contenedor de muestras que contiene un polvo granulado en partículas finas. En el contenedor se deposita una mezcla que contiene polvo de la muestra a analizar y usualmente polvo de Bromuro de Potasio (*KBr*) como medio de dispersión. El *KBr* permite que la radiación se propague libremente entre las partículas de la muestra. Para obtener una buena señal, se recomienda que la proporción entre la muestra a analizar y *KBr* sea no mayor de 1:10. En reflectancia difusa, el haz de radiación se propaga en direcciones indeterminadas dentro del contenedor (Figura 2.6C) y eventualmente es reflejada en múltiples ángulos hacia un elemento llamado colector parabólico compuesto, el cual reúne todas las reflectancias reflejadas y las dirige hacia el detector.
- **Espectroscopia de ruptura inducida por láser LIBS:** Durante el análisis LIBS, un láser de pulsación es orientado hacia la muestra removiendo una

pequeña cantidad de esta a través de miles de pulsos laser por segundo (Figura 2.6D). De esta manera, los átomos en la muestra elevan su temperatura a unos 10,000 °C, provocando un estado de plasma. Cuando el láser toca la muestra, los electrones de las orbitas externas de los átomos son expulsados, lo que provoca inestabilidad. Cuando la pulsación termina el plasma se enfría y los electrones vuelven a su lugar, el exceso de energía liberada cuando el electrón se mueve de una órbita a otra es emitida en forma de radiación en diversas longitudes de onda. Esta radiación es recolectada a través de un cable de fibra óptica para ser procesada por un espectrómetro y obtener las firmas espectrales de la muestra. Debido a las características instrumentales de LIBS, se puede lograr un análisis avanzado que supera las limitaciones de las técnicas analíticas químicas, incluido el tratamiento de muestras pequeñas o no preparativas, el análisis en tiempo real, la aplicación en el sitio y la detección remota de materiales peligrosos (Legnaioli *et al.*, 2014). LIBS puede obtener espectros de suelo que contienen abundante información sobre sus elementos y su textura (Yan *et al.*, 2018).

2.3.3.3 Equipos de laboratorio para análisis espectral de suelos

Con el desarrollo tecnológico en las últimas décadas, se han desarrollado equipos espectrométricos de alta precisión y sensibilidad que permiten determinar la intensidad de radiación en una amplia gama de longitudes de onda del espectro electromagnético. Estos avances han revolucionado la manera en que se estudian las propiedades del suelo, permitiendo una recopilación de datos más detallada y precisa. Hoy en día, los equipos más utilizados para obtener las firmas espectrales del suelo emplean alguna de las técnicas descritas previamente y son capaces de alcanzar resoluciones espectrales menores a 1 nm, lo que brinda una capacidad sin precedentes para analizar con detalle las características del suelo y su composición. En el Cuadro 2.2 se presentan algunos modelos específicos de estos equipos que han sido empleados en investigaciones científicas para el estudio del suelo, destacando su importancia

en el avance de la ciencia de suelos. Cuadro 2.2. Espectrómetros IR para propiedades del suelo.

Modelo	Rango espectral	Fabricante	Lugar de fabricación
MATRIX-I FT-NIR Analyzer	800 - 2564 nm	Bruker Optics	Alemania
NIR Optical Spectrum	900 - 1700 nm	Wuling Company	China
Infrared Optical Spectrum Instrument	900 - 1700 nm	Isuzu Optics Corp.	China
FIELDSPEC4	350 - 2500 nm	Analytical Spectral Devices Inc.	Reino Unido
MATRIXI Fourier Transform Near Infrared Spectrometer	781 - 2779 nm	Bruker	E.U.A.
MobiLIBS system	532 nm con pulsación láser de 5 ns.	IVEA	Francia
QE65000 Spectrometer	200–1100 nm	Ocean Optics	E.U.A.

2.3.3.4. Espectros de suelos

Una vez obtenidos, los espectros del suelo representan la interacción de cada uno de sus componentes fisicoquímicos con la radiación electromagnética. La interacción con estas propiedades se da en forma de absorción, lo que implica una diferencia entre los flujos de radiación (I) y de irradiancia (L), generando patrones o características específicas en cada espectro (Ben-dor & Epema, 2014). En la práctica, para determinada λ estos flujos son la intensidad de radiación de la fuente (I) y la radiación no absorbida por una muestra de suelo (L). Entonces, cuando se obtiene un espectro de absorbancia, implícitamente se obtiene también el espectro de reflectancia, cuya relación está dada por:

$$Abs_{\lambda} = \log_{10} \left(\frac{1}{R_{\lambda}} \right) \quad (2.8)$$

Donde Abs y R son los valores de absorción y reflectancia respectivamente en la longitud de onda λ .

De acuerdo con Ben Dor *et al.*, (2015) cada especie química posee características espectrales únicas. En el suelo, cualquier compuesto orgánico o inorgánico con enlaces covalentes, reacciona a varias frecuencias de radiación en las regiones infrarroja y visible (Donald *et al.*, 2017). En la región infrarroja, la mayoría de las moléculas presentes responden a radiación con frecuencias de 750-25000 nm, es decir en las sub regiones NIR y MIR (Papousek y Aliev, 1982). El manejo de espectros NIR es complejo debido a que los movimientos de los EV no son armónicos (Shao & He, 2011). En contraste, en la región MIR, los espectros tienen una alta absorptividad molar y las características son más específicas. Sin embargo, la reproducibilidad espectral, incluida la relación señal/ruido, de MIRS es más pobre que la de NIRS (Soriano-Disla *et al.*, 2014). Particularmente para la detección de los macronutrientes del suelo NPK, los espectros VIS-NIR han demostrado ser viables en múltiples trabajos (MacAbiog *et al.*, 2020; Masrie *et al.*, 2018; Patel *et al.*, 2020; Qi *et al.*, 2018). Por lo tanto, el uso de la espectroscopia VIS-NIR en suelos se ha logrado posicionar como punta de lanza para su caracterización remota.

2.3.3.5 Factores que alteran la reflectancia en suelos

Además de los efectos de dispersión, en el suelo existen otros factores que modifican la respuesta espectral. Los más importantes son; el contenido de humedad ya que el agua absorbe gran cantidad de radiación; el contenido de óxido de hierro que también absorbe energía en gran parte del espectro VIS-NIR, además de la rugosidad y aspereza causadas por las prácticas de labranza (Warren & Whitehead, 1988). Además, los elementos con comportamiento más dinámico en el suelo como CaO, K₂O, SiO₂, Mn y P₂O₅ pueden tener mayor influencia en la relación de respuesta espectral y fertilidad (Andrade *et al.*, 2020). La temperatura de secado también tiene una influencia notable en la detección en suelos por sensores NIR de acuerdo con lo reportado por Nie *et al.* (2018). Para disminuir el impacto de estos factores en la respuesta espectral de NPK y otros elementos, se realizan diversas preparaciones a las muestras de suelo antes del análisis espectral. El efecto experimental de preparar previamente las muestras mejora la confiabilidad y precisión en la detección en suelos (P. C. Nie

et al., 2017). Los tratamientos típicos a las muestras comprenden secados al aire para disminuir el efecto de la humedad. Así mismo se realizan tamizados y/o moliendas al suelo para homogeneizar el tamaño de partícula (Lourembam, *et al.*, 2017 ; Xiong, *et al.*, 2019) o enfriados (Yan, *et al.*, 2018).

Estas preparaciones han mejorado la precisión de las estimaciones en suelos, Coutinho *et al.*, (2019), destaca el buen desempeño de modelos de predicción de P mediante espectros Vis-NIR y MIR cuando la muestra de suelo se secó a 105 ° C durante 24 h y se tamizó en una malla de 0.25 mm. Por su parte, He *et al.* (2017) reportaron que el mejor tiempo de secado fue de 3 horas con un contenido de agua del 1.03%, para detección de nitrógeno con espectros NIR.

Los resultados experimentales de Nie *et al.*, (2017) mostraron que cuando el contenido de agua era menor y las partículas del suelo eran más pequeñas, fue mejor la precisión y efectividad de la detección de N en suelos utilizando sensores NIR. Esto es respaldado por estudios como los de Xiao & He (2019), donde se encontró que un tamaño de las partículas entre 0,18 y 0,28 mm, mejoró la precisión de predicción de nitrógeno. Por su parte, Yan, *et al.*, (2018) coinciden en que la calidad espectral de nitrógeno mejoró cuando las muestras redujeron y homogeneizaron su tamaño de grano.

La generalización de estos estudios tiene poco alcance, debido a que cada experimento se ha realizado bajo una amplia variación de condiciones geográficas, geológicas, ambientales y de suelo. Por lo tanto, es necesario adaptar la preparación adecuada para cada suelo, condicionado a sus características, el equipo y técnica espectral, así como los objetivos de estudio.

2.4 APRENDIZAJE AUTOMATICO PARA ANALISIS ESPECTRAL EN SUELOS

2.4.1 Aprendizaje automático, generalidades

La inteligencia puede definirse como la capacidad de un sistema para adaptar su comportamiento para cumplir sus objetivos en una variedad de entornos, entendiendo como entorno al contexto, las condicionantes y los límites del problema. Para una máquina, tanto los objetivos como los entornos están proporcionados por una persona, por lo tanto, el concepto de inteligencia artificial

(IA) tomaría mayor sentido en un paradigma de extensión y potenciador de la inteligencia humana (Domingos, 2015). Es un campo de estudio multidisciplinario que se centra en el desarrollo de sistemas y programas capaces de simular y replicar procesos de razonamiento humano, como el aprendizaje, la lógica y la toma de decisiones a través de la teoría matemática y estadística, para resolver problemas complejos. Dentro de la IA existen diversas técnicas y enfoques, como el aprendizaje automático (ML por Aprendizaje automático), el procesamiento del lenguaje natural, la visión por computadora, los sistemas expertos, la lógica difusa, entre otros. Estas técnicas pueden complementarse entre sí y utilizarse en combinación para abordar problemas complejos en diferentes dominios.

ML que se enfoca en el desarrollo de modelos a partir de algoritmos. Estos modelos permiten a las máquinas aprender y mejorar automáticamente a partir de datos, aprendiendo y adaptándose al entorno en que se opera. En ML se estudian métodos capaces de imitar, comprender y ayudar en las tareas de procesamiento de información. A menudo, estos métodos no están dirigidos a imitar el procesamiento humano, sino a mejorarlo (Barber, 2020).

2.4.2 Conceptos básicos en aprendizaje automático

Dentro de ML están implícitos conceptos fundamentales que permiten una comprensión más amplia, a continuación, se describen brevemente algunos de los más importantes.

Algoritmos

Es una secuencia ordenada de instrucciones precisas y acotadas, que indican a una computadora los pasos a seguir para procesar y analizar datos, para alcanzar un objetivo. De esta forma las máquinas pueden aprender de los datos y tomar decisiones utilizando una amplia gama de estructuras y lenguajes (Domingos, 2015). Algunas estructuras básicas fundamentales comprenden; estructuras secuenciales, condicionales y de iteración. Estas indican el orden en que las instrucciones son ejecutadas dentro de los diferentes entornos de programación como Python, Matlab, Java, R, C++, etc.

Aprendizaje

Se refiere al desarrollo de conocimiento (modelos) y habilidades (tareas) que una maquina puede obtener a partir de diversos procesos de análisis de datos generados por los algoritmos. Murphy, (2012) clasifica el aprendizaje de máquina en tres grandes campos de desarrollo, los cuales corresponden a los objetivos básicos y la forma de aprendizaje que emplean, el aprendizaje supervisado, el no supervisado y el aprendizaje reforzado (Cuadro 2.3). Para la resolución del problema de detección en suelos, como en otros problemas complejos, se pueden combinar las tres formas de aprendizaje, siendo el aprendizaje supervisado el principal enfoque (Liakos et al., 2018).

El aprendizaje supervisado consiste en aprender de un conjunto de datos con ejemplos etiquetados proporcionados por un supervisor externo informado. Cada ejemplo es una descripción de una situación junto con una especificación o etiqueta de la acción correcta que el sistema debe tomar para esa situación (Sutton & Barto, 2019). Parte de un conjunto de datos D etiquetados:

$$D = \{(x_i, y_i), (x_{i+1}, y_{i+1}), \dots, (x_n, y_n)\} \quad , i = 1:n$$

En el conjunto D los datos se organizan en pares, x_i representa las entradas del modelo e y_i las salidas o etiquetas. El conjunto de n datos de entrada se puede denotar como X y el conjunto de n etiquetas como Y . El objetivo de este aprendizaje es encontrar y modelar una función de distribución subyacente en los datos de X , que permita relacionarlos con los de Y , es decir: $Y = f(X)$.

Cuadro 2.3. Formas de aprendizaje en ML

Aprendizaje	Datos	Tarea	Objetivo
Supervisado	Etiquetados $D = \{x_i, y_i\}_{i=1}^N$	Clasificación	Aprender a mapear las salidas y a partir de una función $f(X)$, con $y \in \{1, \dots, C\}$
		Regresión	Aprender a mapear las salidas y a partir de una función $f(X)$, con $y \in \mathbb{R}$
No Supervisado	No etiquetados	Agrupamiento	Agrupar de datos con características similares

	$D = \{x_i\}_{i=1}^N$	Reducción de dimensionalidad	Identificar factores latentes de datos.
		Estructuración grafica	Obtener estimadores de densidad de probabilidad conjunta.
		Imputación	Inferir valores para variables con mediciones incompletas.
			Tomar decisiones secuenciales para maximizar una recompensa acumulada en un entorno determinado
Reforzado	No etiquetados $D = \{x_i\}_{i=1}^N$	Control y Optimización	

En el aprendizaje no supervisado o descriptivo solo existen entradas para los algoritmos, $D = \{x_i\}_{i=1}^N$, y el objetivo es encontrar "patrones" en los datos. Está basado en la estimación de densidad θ , lo cual se realiza construyendo modelos de la forma $p(x_i|\theta)$. Aquí la densidad de los datos no condiciona el aprendizaje, sino que es el elemento de estudio.

El aprendizaje reforzado por otra parte se enfoca en el desarrollo de algoritmos y sistemas que aprenden a través de la interacción directa con un entorno dinámico. A diferencia del aprendizaje supervisado, donde los datos de entrenamiento están etiquetados con respuestas correctas, y el aprendizaje no supervisado, donde se busca encontrar patrones y estructuras en los datos sin etiquetas, el RL se basa en la idea de que un agente aprende a través de ensayo y error. Es ideal para problemas iterativos donde es necesario aprender de la experiencia (Sutton & Barto, 2019).

Entrenamiento sobreajuste y desajuste

El entrenamiento del modelo implica un proceso de ajuste de los factores que definen a f , a través de un subconjunto $D_1 \in D$ al cual se conoce como conjunto de entrenamiento o calibración. Generalmente se maneja que D_1 contenga el 75% de los datos en D . El objetivo es encontrar los parámetros o coeficientes óptimos a través de algoritmos que minimicen la diferencia entre los valores de salida generados por los modelos (Y') y los valores reales de la variable objetivo

(Y). Este proceso es conocido como entrenamiento o calibración del modelo y depende en gran medida de la capacidad del modelo.

La capacidad de aprendizaje de un modelo está dada por la estructura y los parámetros que lo conforman, esto se refleja en su habilidad para aprender distribuciones de diferentes grados de complejidad. Un modelo con alta capacidad puede aprender patrones de distribución más complicados.

Por otro lado, el sobreajuste y el desajuste son problemas de rendimiento del modelo, ocurren cuando un modelo tiene una capacidad excesiva o insuficiente respectivamente de relacionar los datos. En este sentido, el objetivo del aprendizaje, es encontrar un equilibrio entre la capacidad del modelo y la complejidad de los datos para lograr un rendimiento óptimo y una buena generalización en la predicción (Murphy, 2012).

Predicción

Una vez que el modelo es entrenado debe ser probado en otro conjunto de datos $D_2 \in D$, siendo D_2 un subconjunto de D que contiene el resto de los datos. El objetivo es conocer el rendimiento de los modelos ante datos no conocidos, a través de diversas métricas de evaluación, algunas de las más utilizadas en ciencias de suelos se presentan en el Cuadro 2.4. La elección de la métrica adecuada depende de los objetivos específicos del problema y las características de los datos. En general, la raíz del Error Cuadrado Medio (RMSE) y el Coeficiente de Determinación (R^2) son métricas ampliamente utilizadas en la evaluación de modelos de predicción en diversos campos. Sin embargo, en el contexto de modelos espectrales de suelos, la Relación de Rendimiento y Desviación (RPD) y el Relación entre el Rendimiento y la Distancia Intercuartílica (RPIQ) pueden ser especialmente relevantes, ya que están diseñados para evaluar la precisión de las predicciones en relación con la variabilidad y la dispersión de los datos espectrales. Generalmente, para ambos se adopta un umbral de decisión donde valores superiores a 2 indican buenos modelos para aplicaciones agrícolas (Bellon-Maurel *et al.*, 2010).

Tareas

En espectroscopia de suelos, se busca relacionar variables objetivo (Y) en función de características espectrales (X) en forma de longitud de onda. Cuando los valores de Y son reales y continuos, el objetivo es cuantificar a Y . Si las etiquetas corresponden a valores discretos o categóricos el objetivo es clasificarlas. Por lo tanto, para detección en suelos, las tareas principales son la regresión y la clasificación o reconocimiento de patrones.

Cuadro2.4. Métricas de evaluación de modelos de regresión en ML

Métrica	Definición	Descripción
Error cuadrado medio (MSE)	$\frac{1}{N} \sum (Y - Y')^2$	Es el promedio de los errores al cuadrado entre las N predicciones Y' y los valores reales Y
RMSE	$\sqrt{MSE}$	Raíz del error cuadrado medio
Coeficiente de Determinación R^2	$1 - \left(\frac{\sum (Y - Y')^2}{\sum (Y - \bar{Y})^2} \right)$	Mide la proporción de la varianza total de los datos que es explicada por el modelo, varía entre 0 y 1.
RPD	$\frac{DE}{\sqrt{MSE}}$	Es la desviación estándar (DE) de los datos etiquetados de prueba, sobre el RMSE.
RPIQ	$\frac{IQ}{\sqrt{MSE}}$	Es la distancia Inter cuartil $IQ = (Q3 - Q1)$ sobre el RMSE.

Clasificación: En un modelo de clasificación, se busca asignar instancias de X a una o varias categorías o clases discretas en Y . Es decir $Y \in \{c_1, c_2, \dots, C\}$, siendo C el número de clases c_i . El modelo de clasificación estima la probabilidad de pertenencia a cada clase y utiliza un criterio de decisión (por ejemplo, un umbral) para asignar la instancia a una clase específica. Por lo tanto, un modelo de clasificación se puede expresar mediante una función de probabilidad condicional de la siguiente manera:

$$f(X) = P(Y \in c_i | X) + \varepsilon$$

Esta es la probabilidad condicional de que la variable de salida Y pertenezca a la clase c_i dados los valores de entrada X y es la base de los clasificadores. Aquí ε representa el término de error o ruido, que captura la variabilidad no explicada por el modelo (Murphy, 2012).

Regresión: Los modelos de regresión suponen una relación subyacente entre X e Y . En espectroscopia de suelos, la naturaleza de esta relación es uno de los objetivos de estudio principales. Teniendo en cuenta la ley teórica de Beer-Lambert que se encuentra detrás de NIRS, por lo general se suponía una función f lineal, lo que ha contribuido en gran medida al desarrollo de métodos de calibración multivariados lineales adaptados a datos espectrales (Martens *et al.*, 2003). Sin embargo, dependiendo de las condiciones del suelo, la cantidad de datos disponibles y otros delimitantes del problema pueden resultar útiles diferentes paradigmas (Ahmadi *et al.*, 2021; Barra *et al.*, 2021). Algunos de los enfoques más comunes son:

- Regresión lineal. Son modelos de calibración que tienen la forma:

$$f(X) = \beta^0 + \beta^1 X^1 + \beta^2 X^2 + \dots + \beta_p X_p + \varepsilon$$

Donde $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ son los coeficientes que se estiman a partir de las características espectrales de entrada X .

- Redes neuronales (no lineal). La notación básica de estos modelos es la siguiente:

$$f(X) = \sigma(w_0 + w_1 X_1 + w_2 X_2 + \dots + w_p X_p)$$

Donde σ es una función de activación, la cual puede ser de diferentes tipos, siendo algunos de los más comunes, las funciones: sigmoide (Sigmoid), ReLU (Rectified Linear Unit) o la tangente hiperbólica (tanh). Su propósito principal es introducir la no linealidad, lo que le permite aprender relaciones y patrones más complejos. Por otro lado, $w_0, w_1, w_2, \dots, w_p$ son los pesos que se ajustan durante el entrenamiento.

- Árboles de decisión: están basados en la división automática del espacio de características X en M regiones R_m . Esta división se realiza a partir de variables de decisión y puntos de corte llamados nodos. Cada región está asociada con un valor de Y como se indica:

$$f(X) = \sum_m^M C_m | (X \in R_m)$$

Donde C_m es el promedio de la variable dependiente Y , cuando las variables de decisión toman valores correspondientes a determinada R_m

- Máquinas de vectores de soporte SVM: Para regresión, los modelos SVM buscan ajustar los datos en X a una función lineal de manera óptima (V. N. Vapnik, 1999). Para lo cual utilizan un subespacio de características determinado por una función núcleo (kernel). El modelo SVM en el subespacio se define de la siguiente manera:

$$f(X) = W^T \varphi(X) + b$$

Donde, W^T es un vector de pesos que se ajustan en el entrenamiento, b es el sesgo y $\varphi(X)$ es la función núcleo. Para espectroscopia de suelos se conocen diversos núcleos útiles como el Sigmoidal o el Polinomial (Ji-yong *et al.*, 2013), sin embargo uno de los más utilizados es la función de base radial (RBF) que se define como sigue:

$$\varphi(X) = RBF = e^{-\frac{(\|x-x'\|^2)}{2\sigma^2}}$$

Donde $\|x - x'\|$ es la distancia euclidiana entre los vectores de características x y x' . σ^2 es un parámetro de escala que determina la amplitud o el alcance del mapeo no lineal en el subespacio RBF.

En SVM existe otro parámetro de interés γ , el cual es un parámetro de regulación que está relacionado con la complejidad del modelo y cómo se adapta a los datos. Este y el parámetro de escala son hiper-parámetros a optimizar para usar SVM con RBF. En espectroscopia de suelos, diversos autores (Keyvan *et al.*, 2021; Wenjun *et al.*, 2014; S. Xu *et al.*, 2018) utilizan un método de validación cruzada para el ajuste.

2.4.3 Subcampos de aprendizaje automático para suelos

A partir del ML se derivan diferentes metodologías de aprendizaje y análisis de datos, las cuales emplean diversas formas de aprendizaje, acoplan diferentes

técnicas de regresión, clasificación, optimización, etc. y se enfocan en la resolución de una amplia gama de problemas en ingeniería. En el Cuadro 2.5 se proporciona una visión general de algunos subcampos del ML, sus enfoques, aprendizaje predominante, tareas comunes, objetivos generales y ejemplos de algoritmos representativos. Particularmente, el aprendizaje profundo (DL) ha tenido un impacto muy importante en el estudio y manejo de diversas tareas agrícolas incluyendo el manejo general de suelos (Liakos *et al.*, 2018). DL se basa en redes neuronales profundas que están compuestas por múltiples capas de unidades interconectadas. Estas redes se caracterizan por su arquitectura y su función de activación. Además, son entrenadas con diversas técnicas para ajustar los pesos de las conexiones, para lo cual utilizan grandes conjuntos de datos etiquetados. Algunas de las más utilizadas en análisis de suelos y el sector agrícola se indican a continuación:

- **Redes Neuronales de Retropropagación (BPNN).** Son un tipo de red neuronal que utiliza el algoritmo de retropropagación para entrenar y ajustar los pesos de las conexiones entre las neuronas (Rojas, 1996). El proceso de retropropagación implica dos fases principales: propagación hacia adelante y propagación hacia atrás. Durante la propagación hacia adelante, los datos de entrada se introducen en la red y se propagan a través de las capas ocultas hasta llegar a la capa de salida, generando una predicción. Luego, se calcula el error entre la predicción generada y el valor real deseado. En la fase de propagación hacia atrás, el error calculado se propaga en sentido contrario, desde la capa de salida hasta las capas ocultas. Utilizando la regla de la cadena y el cálculo del gradiente, se ajustan los pesos de las conexiones en función del error, con el objetivo de minimizarlo. Estos algoritmos son los más utilizados en análisis espectral de suelos (Song *et al.*, 2019; S. Xu *et al.*, 2018).
- **Redes Neuronales Feedforward (FNN).** Estas redes tienen una estructura secuencial con capas de neuronas que transmiten información hacia adelante. Sus funciones de activación comunes son la sigmoide, ReLU, y Tanh. Es común el uso de versiones de este tipo de redes como

la red de aprendizaje extremo (ELM), la cual ha sido probada para analizar diversos elementos químicos en el suelo (Sadgrove *et al.*, 2021; R.-J. Vestergaard *et al.*, 2021; M. Yang *et al.*, 2019). ELM es una red de alimentación directa de una sola capa oculta generalizada (SLFN) con pesos asignados aleatoriamente en su única capa oculta y un peso en la capa de salida que se calcula directamente por los mínimos cuadrados método. Todo el proceso de aprendizaje se completa en una ronda, sin necesidad de iteraciones; por lo tanto, este algoritmo funciona a una velocidad de aprendizaje extremadamente rápida (Huang *et al.*, 2006).

- **Redes Neuronales Convolucionales (CNN).** Las CNN están diseñadas específicamente para el procesamiento de datos estructurados, como imágenes. Utilizan la función de activación “ReLU” con capas convolucionales y de agrupación para extraer características y reducir la dimensionalidad (Albawi & Mohammed, 2017). En suelos han sido utilizadas para análisis de pH, textura y carbono (Wartini *et al.*, 2019).
- **Redes Neuronales Recurrentes (RNN).** Las RNN tienen conexiones retroalimentadas que les permiten procesar secuencias de datos. Pueden mantener y utilizar información de estados anteriores, lo que las hace adecuadas para tareas de procesamiento de lenguaje y otras aplicaciones que involucran datos secuenciales. Sus funciones de activación comunes son: “sigmoide”, “tanh”, “ReLU”, “GRU” (Salehinejad *et al.*, 2018). RNN han sido utilizadas para sustituir los métodos comunes de análisis de suelo por espectroscopia y ML (J. Yang *et al.*, 2020)
- **Redes Neuronales Generativas Adversariales (GAN).** Las GAN son un sistema de dos redes neuronales artificiales, una de generación y otra de discriminación que compiten entre sí. El generador crea muestras similares a los datos de entrenamiento, mientras que el discriminador distingue entre muestras reales y falsas (Goodfellow *et al.*, 2020). En análisis espectral de suelos, las GAN han sido utilizadas para generar espectros sintéticos, con el objetivo de robustecer modelos entrenados con bases de datos poco numerosas (Jiang *et al.*, 2023).

El aprendizaje Reforzado (RL) por otro lado, se basa en la interacción de un agente de aprendizaje (máquina o software) con un entorno. RL es aprender a tomar decisiones en base a un estado actual y recibir retroalimentación para maximizar una señal de recompensa numérica. Al agente no se le dice qué acciones tomar, él debe descubrir qué acciones producen la mayor recompensa al probarlas. Estas dos características, la búsqueda por ensayo y error y la recompensa diferida, son las dos características distintivas más importantes del aprendizaje por refuerzo. Los algoritmos de RL pueden utilizar métodos de aprendizaje supervisado y no supervisado dentro de su proceso de aprendizaje (Sutton & Barto, 2019). Así mismo, existen otros subcampos de ML con diversos enfoques, con características y objetivos muy particulares. Los métodos de aprendizaje explicativo (Explainable Learning) por ejemplo, permiten comprender cómo y por qué se toman ciertas decisiones y proporcionan justificaciones basadas en características relevantes de los datos.

El aprendizaje de transferencia (Transfer Learning) utiliza conocimientos adquiridos en una tarea fuente para mejorar el rendimiento y transferir el aprendizaje a una tarea objetivo que se relacione (Lanfa Liu et al., 2018).

Otros como el aprendizaje activo (Active Learning) interactúan con un "oráculo" o experto humano, al cual solicitan etiquetas adicionales en momentos estratégicos durante el entrenamiento, con el objetivo de desarrollar modelos que puedan seleccionar de manera inteligente ejemplos informativos o desafiantes para etiquetar y mejorar el rendimiento (Brame, 2007).

Cuadro 2.5. Subcampos del Aprendizaje automático

Sub-camp o	Datos	Tareas	Objetivo	Algoritmos	Aplicaciones
---------------	-------	--------	----------	------------	--------------

Aprendizaje profundo	Sv, NSv	Clasificación, Regresión, Segmentación, Generación, etc	Aprender representaciones de datos complejos y de alto nivel	Redes Neuronales Convolucionales (CNN), Redes Neuronales Recurrentes (RNN), Redes Generativas Adversariales (GAN)	Reconocimiento de imágenes, Lenguaje Natural, Vehículos autónomos, Generación de datos.
Aprendizaje reforzado	Reforzado	Toma de decisiones secuenciales, control	Aprender a través de la interacción con el entorno	Q-Learning, Algoritmo Genético	Control, Robótica, Optimización
Aprendizaje activo	Sv, NSv	Clasificación, Regresión	Mejorar el rendimiento del modelo con un menor costo de etiquetado.	SVM activo, Query by Committee	Etiquetado de datos, Detección de fraudes
Aprendizaje transferido	Sv	Clasificación, Regresión	Transferir conocimientos de una tarea fuente a una tarea objetivo	Redes Neuronales Pre-entrenadas, Fine-tuning	Reconocimiento de objetos, Traducción automática
Aprendizaje explicativo	Sv, NSv	Interpretación, Explicación	Proporcionar explicaciones claras y comprensibles sobre el proceso de toma de decisiones del modelo	Regresión LASSO, Reglas de decisión	Sistemas de recomendación, Análisis médico
Aprendizaje de ensamble	Sv, NSv	Clasificación, Regresión, Clustering	Mejorar la precisión y la robustez de los modelos mediante la combinación de múltiples modelos base	Random forest, Gradient boosting, K-means	Predicción de precios, segmentación de clientes,

Por otro lado, el aprendizaje de ensamble (Ensemble Learning) también tiene como objetivo principal de mejorar la generalización, esto se logra reduciendo el sesgo y la varianza a partir de métodos de combinación, como votación, promedio

o ponderación, para generar una predicción final basada en las predicciones individuales de modelos base (Dong *et al.*, 2020).

2.4.4 Desarrollo de modelos espectrales con aprendizaje automático para predicción de nutrientes de suelo

En este trabajo, el proceso de análisis espectrométrico en suelos mediante Aprendizaje Automático consta de tres etapas esenciales: obtención de datos, selección de variables y predicción de nuevos datos. La Figura 2.7 proporciona una representación visual de la secuencia ordenada de cada uno de estos procedimientos.

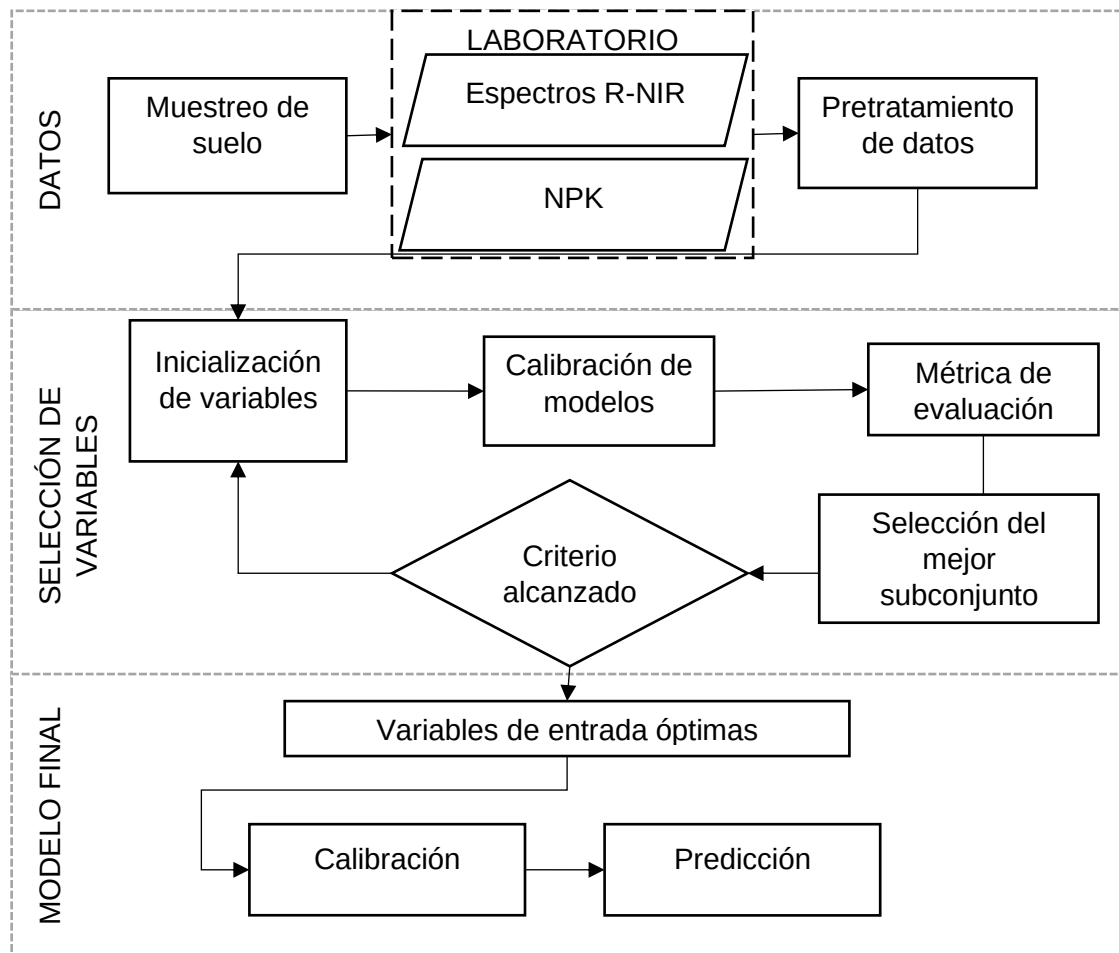


Figura 2.7. Método de análisis espectral de suelos con ML

Creación de bases de datos espectrales de suelo

En esta etapa se incluye el muestreo en campo y la preparación de las muestras para poder realizar análisis espectrométricos y de fertilidad (NPK).

Muestreo

El muestreo de suelo constituye el primer paso del análisis químico y físico. El cuidado que se tenga en la toma de la muestra del suelo determinará en gran parte la validez de los resultados. En esta etapa, es necesario determinar factores que permitan realizar un muestreo representativo del sitio de estudio. De acuerdo con el INIFAP estos factores son:

- Época de muestreo. Se recomienda realizar el muestreo después de haber preparado el terreno para el siguiente ciclo agrícola, entre 1 a 2 meses antes de la siembra. En el caso de un suelo salino el muestreo puede realizarse en cualquier época del año.
- Forma de muestreo. Se determina en función de la variación a lo largo ancho y profundo del terreno. Para obtener muestras representativas, el área deberá ser lo más homogénea posible. Algunos de los métodos más utilizados por su eficiencia son; el método de Zigzag consiste en obtener de 7 a 20 submuestras por cada 5 a 10 hectáreas de terreno, siguiendo una ruta que forme ángulos alternativos entrantes y salientes (Figura 2.8A), en superficies mayores se debe subdividir en áreas uniformes para la toma de las submuestras. Si el terreno presenta heterogeneidad por algunas variables como color, textura, pendiente, fertilidad, etc., se puede realizar un muestreo estratificado, en el cual se agrupan las zonas homogéneas y después se determina una ruta de muestreo en Zigzag (Figura 2.8B). Para realizar análisis espectrales, es importante tomar en cuenta la textura del suelo como factor de estratificación, ya que esta es determinante en la reflectancia del suelo (Ben-dor & Epema, 2014). Si el terreno es uniforme, se puede realizar el muestreo por el método de cinco de oros (Figura 2.8C).

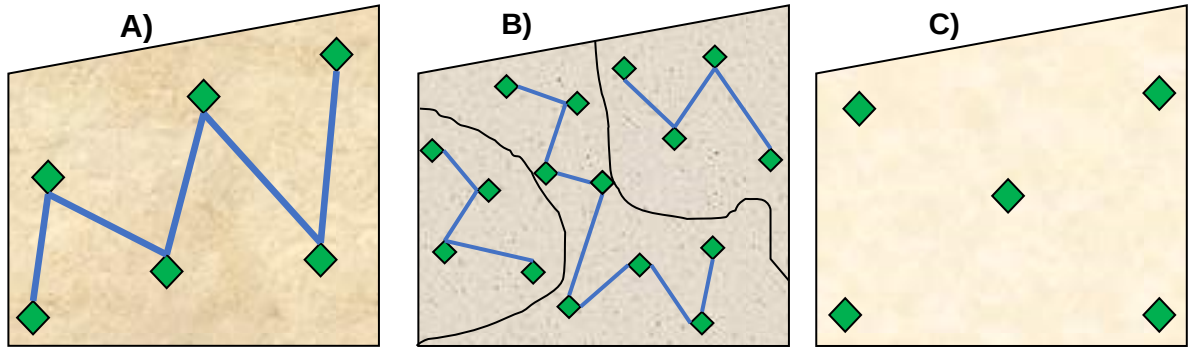


Figura 2.8 Muestreo en zigzag (A), estratificado(B), cinco de oros(C).

- Localización. En un marco de AP, el sitio de estudio debe ser debidamente localizado y registrado. De manera que sea posible mapear la variabilidad del terreno de manera precisa, así como para la ubicación de la zona en futuros estudios. La posición del sitio y de los puntos de muestreo puede realizarse a través de un sistema de posicionamiento global (GPS), se recomienda una precisión de al menos 2 m.
- Profundidad de muestreo. Para propósitos de estudios de fertilidad del suelo, se recomienda realizar el muestro a una profundidad de influencia radical de las plantas. Para cultivos anuales, el muestreo se efectúa sobre la capa superficial del suelo, entre 25 y 30 cm de profundidad. Para frutales y especies forestales el muestreo se realiza a los 40 cm.
- Obtención y tamaño de la muestra. Se recomienda utilizar una pala de acero inoxidable o una barrena para perforar la superficie del terreno y extraer muestras de suelo a una profundidad adecuada. El tamaño de la muestra dependerá de los métodos de análisis utilizados, generalmente entre 300 y 500 gr es suficiente para los métodos comunes en laboratorios. Una vez obtenida la muestra debe ser debidamente almacenada e identificada en un envase plástico que no permita su contaminación.

Análisis de NPK en laboratorio

Los métodos de análisis de laboratorio para NPK permiten determinar los niveles de estos elementos químicos. A medida que los datos de contenido de NPK se acumulan, se convierten en una valiosa base de datos que se puede aprovechar

para desarrollar modelos predictivos de aprendizaje supervisado. A continuación, se mencionan algunos métodos estándar.

Nitrógeno: El método estándar para la determinación en laboratorio de Nitrógeno en suelos es el procedimiento de Kjeldahl (Sáez-Plaza *et al.*, 2019). Es un índice que indica la disponibilidad de nitrógeno para las plantas en forma de nitratos NO_3^- y amonio NH_4^+ . El método se engloba en la categoría de medios por digestión húmeda. En el desarrollo de modelos de espectrales para fertilidad de suelos con aprendizaje supervisado, ha sido ampliamente utilizado a nivel global, así lo indican estudios previos (Ma *et al.*, 2019; Mohamed *et al.*, 2020; Rodríguez-Pérez *et al.*, 2021; M. Yang *et al.*, 2020; etc.).

Fósforo: La determinación de P aprovechable para suelos se realiza a través del método de Olsen (Horta & Torrent, 2007) o de Bray (Bray & Kurtz, 1945) según la acidez del suelo. El método de Olsen es utilizado en estudios de fertilidad de suelos alcalinos, así como para suelos neutros. El P es extraído del suelo con una solución de NaHCO_3 0.5 M ajustada a un pH de 8.5. En suelos neutros, calcáreos o alcalinos, conteniendo fosfatos de calcio, este extractante disminuye su concentración de Ca en la solución a través de una precipitación del CaCO_3 , por lo tanto, la concentración de P en la solución se incrementa. La determinación de P en suelos neutros y ácidos se realiza mediante el método de Bray y Kurtz. En suelos ácidos, los cuales contienen fosfatos de Al y Fe, la concentración de P se incrementa conforme el pH se eleva. Para la determinación, la solución extractora consiste en una combinación de HCl y NH_4F la cual remueve las formas de P ácido solubles (fosfato de calcio, aluminio y hierro). El NH_4F disuelve los fosfatos de Al y Fe al formar un ion complejo con estos iones metálicos en la solución.

Potasio: La determinación de K se realiza a partir del método AS-19 de la norma para estudio, muestreo y análisis de fertilidad de suelos (NOM-021-RECNAT-2000). Donde se determinan los cationes solubles en muestras de suelo. La extracción se realiza a través de espectrofotometría de emisión atómica en los extractos diluidos, con ayuda de un espectrofotómetro de flama.

Obtención de espectros NIR en laboratorio

Preparación de muestra: La obtención de espectros R-NIR de suelos es precedida de una etapa de preparación de la muestra, ya que existen factores que alteran el espectro y dificultan la precisión en los modelos de predicción (Cuadro 2.4). La preparación de muestras de suelo para análisis espectral se describe a continuación;

1. Remoción de rocas y desechos orgánicos.
2. Secado al aire o en horno; Los valores típicos van de 80 a 110° C por periodos de hasta 24 horas.
3. Molienda y Tamizado; Para garantizar la homogeneidad en tamaño de partícula, el suelo debe ser triturado y pasado por un tamiz de ≤ 2 mm (Lin *et al.*, 2017; Yan *et al.*, 2018).

Análisis espectral: En el laboratorio, existen dos tipos de factores principales que afectan el espectro NIR del suelo. Los cuales pueden producir espectros de suelo ruidosos e inconsistentes si no se controlan. Estos factores son clasificados en sistemáticos y no sistemáticos (Ben Dor *et al.*, 2015).

- a) Factores sistemáticos: Son aquellos que surgen de variables controladas, que pueden cambiar de un instrumento a otro. Para evitar efectos sistemáticos, factores como el blanco de referencia, la configuración espectral, la geometría de medición, el operador, la distribución del tamaño de las partículas y las condiciones ambientales de medición, se deben de mantener constantes.
- b) Factores no sistemáticos: Son los que surgen de variables incontrolables, como el ruido aleatorio, la refracción de la muestra o la longitud de onda. Para minimizar los efectos, se deben mantener constantes los factores de instrumentación (fuente de iluminación, salida del detector, etc.), así como la preparación de muestras.

Para la obtención del espectro R-IR de suelo, los pasos son los siguientes:

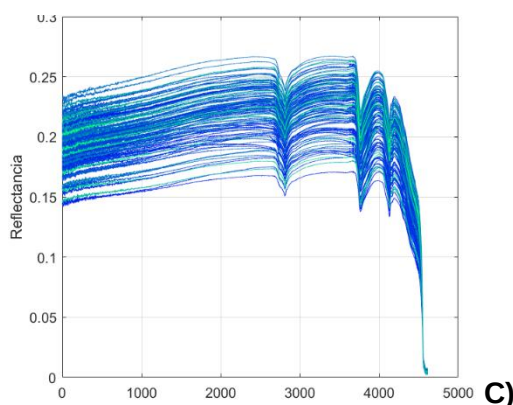
1. Calibración del espectrómetro. Se introduce una muestra estándar blanca en el portaobjetos del equipo (Figura 2.9A). Se debe realizar al inicio de cada jornada y/o en cada lote de muestras.
2. Acondicionamiento de la muestra. Se toman 10 gr de suelo preparado y se vierte en una celda de cuarzo que se introduce en el portaobjetos del espectrómetro una vez calibrado (Figura 2.9B).
3. Configuración del espectrómetro. Los parámetros básicos a configurar son: tipo de análisis (absorción, reflexión o emisión), técnica espectral (IR, NIR, MIR), tipo de muestra (sólida, líquida, tejidos, polvos).
4. Medición espectral. Cuantificación de la interacción (absorbancia, reflectancia, etc.) entre radiación y muestra (Figura 2.9C).



A)



B)



C)

Figura 2.9. Muestra estándar de calibración(A), muestra colocada en el espectrómetro (B) y medición de la respuesta espectral(C).

2.3.4.2 Pretratamiento de los espectros

Como ya se mencionó, en medios difusos como el suelo, generalmente existen diversos factores que pueden provocar alteraciones a la respuesta espectral y aparecer en forma de ruido y otras alteraciones. De manera que las correlaciones con los elementos presentes en la muestra también se pueden modificar, dificultando la calibración de modelos predictivos (Buddenbaum & Steffens, 2012). Por lo que, es importante aplicar correcciones que permitan eliminar estas alteraciones. Estas correcciones son en esencia manipulaciones matemáticas y estadísticas llamadas pretratamientos. Los pretratamientos son aplicados a los datos espectrales antes de realizar el análisis multivariado para mejorar los modelos (Coutinho *et al*, 2019; Yao *et al.*, 2018; Zhang, *et al*, 2016a; Zhang *et al*, 2019). En el Cuadro 2.6 se muestran las técnicas utilizadas en espectroscopia de suelos. Se identifican dos objetivos básicos al aplicar pretratamientos; el suavizado y las correcciones de dispersión.

Suavizado espectral. El suavizado es el proceso de eliminar las variaciones aleatorias que aparecen en forma de ruido de alta frecuencia en un espectro. Tiene un efecto de alisado en la respuesta espectral, pero dependiendo del nivel de suavizado, pueden causar pérdida de información espectral útil.

Los pretratamientos de suavizado más utilizados en espectros VIS-NIR de suelos son; el suavizado de promedio móvil (MAS), el filtro de Savitzky Golay y el análisis wavelet.

- Suavizado de promedio móvil. MAS fue desarrollado en la década de 1920, es el proceso más antiguo para suavizar datos y continúa siendo una herramienta útil en la actualidad. Este método trabaja bajo el supuesto de que las observaciones cercanas en un rango espectral tienen valores similares. El filtro reduce el ruido promediando puntos contiguos para resaltar la señal (Guiñón *et al.*, 2007).

Cuadro 2.6. Pretratamientos para corrección de espectros de suelo.

Corrección	Pretratamiento	Impacto	Ecuación	Referencia
Suavizado	Filtro de Savitzky Golay	Remueve el ruido de alta frecuencia. Tiende a mantener características de la distribución inicial	$\frac{1}{M} \sum_{i=-m}^m C_i Y_{j+i}$	(Qi <i>et al.</i> , 2018)
	Promedio móvil (MAS)	Suaviza el ruido de baja frecuencia	$\frac{1}{M} \left[\left(\sum_{i=1}^{M-1} R_i \right) + R_M \right]$	(He <i>et al.</i> , 2007)
	Análisis Wavelet	Suaviza el espectro conservando sus características críticas	$\sum_{jk \rightarrow \infty} c_{jk} \Psi_{jk}(t)$	(Y. Zhang <i>et al.</i> , 2016)
Dispersión	Eliminación continua	Elimina los efectos de fondo continuo		(Dotto <i>et al.</i> , 2017)
	Corrección de dispersión multiplicativa (MSC)	Mitiga los efectos de la dispersión	$x_i = a_i + b_i X + e_i$ $MSC_i = (x_i - a_i)/b_i$	(Buddenbaum & Steffens, 2012)
	Eliminación de tendencias	Corregir la línea base del espectro, resalta características		(Dotto <i>et al.</i> , 2017)
	Variación normal estándar (SNV)	Realiza tanto el desplazamiento como el escalado del espectro resaltando los máximos y mínimos	$\frac{R - \mu_R}{\sigma_R}$	(H. Yang <i>et al.</i> , 2012)
Derivativo	Savitzky Golay derivativo	Reduce la variación de la línea de base y mejorar las características espectrales, aplanar el espectro y resalta máximos y mínimos.	$\frac{d}{dY} \left[\frac{1}{M} \sum_{i=-m}^m C_i Y_{j+i} \right]$	(Vestergaard <i>et al.</i> , 2021)
	Norris Williams	Resalta características espectrales sutiles	$\frac{(y_{i+1} - y_i)}{x_{i+1} - x_i}$	(Rinnan <i>et al.</i> , 2009)
	GAP	Elimina los efectos aditivos y multiplicativos, elimina efectos de fondo		(Slaughter <i>et al.</i> , 2001)

- Filtro de Savitzky-Golay. Savitzky-Golay (SG), se basa en el cálculo de una regresión polinomial local, que da como resultado una función similar a los datos de entrada, pero suavizada. Su principal ventaja es que tiende a preservar características de la distribución inicial, tales como los máximos y mínimos relativos, así como parámetros de frecuencia (Savitzky & Golay, 1964).
- Análisis Wavelet. Es un método utilizado para retener información efectiva y eliminar ruido de la señal causado por factores no sistemáticos. El análisis busca conservar los valores críticos de la señal espectral y eliminar el ruido aleatorio. Para lo cual es necesario determinar el nivel de descomposición espectral que permita retener la mayor cantidad de información útil. La determinación del nivel óptimo de descomposición se realiza evaluando el indicador de relación relativa señal/ruido (RSNR) (Y. Zhang et al., 2016).

Correcciones de dispersión. La variabilidad física entre muestras difusas como el suelo (tamaño y forma de partícula, superficie de la muestra), provocan variaciones incontroladas en la longitud del camino óptico de la radiación. Dando paso a los efectos de dispersión de luz multiplicativa y aditiva, los cuales pueden escalar el espectro y modificar su línea base (Jin *et al.*, 2012). La línea base de un espectro está determinada por aquellas regiones donde no existen absorciones significativas (idealmente, la intensidad en estas regiones del espectro es de cero unidades de absorbancia ó 100 % de transmitancia).

El grado de dispersión depende de factores no sistemáticos como la longitud de onda y del índice de refracción de la muestra. Debido a estos factores la línea base puede inclinarse, desplazarse o incluso curvarse. Esto podría alterar las variaciones espectrales relacionadas con la composición química de las muestras. Para corregir los efectos de dispersión, se emplean funciones matemáticas que compensan las diferencias en el camino óptico (Moros Portolés, 2007).

Los pretratamientos utilizados generalmente en espectros VIS-NIR de suelos incluyen: normalización (Burger & Geladi, 2007; Rinnan *et al.*, 2009), corrección

de dispersión multiplicativa (MSC) que evidentemente remueve efectos multiplicativos pero también aditivos (Cambule *et al.*, 2012; Lin *et al.*, 2017), eliminación de tendencias (Dotto *et al.*, 2017; Rodríguez-Pérez *et al.*, 2021; R.-J. Vestergaard *et al.*, 2021), la variación normal estándar (SNV) (R.-J. Vestergaard *et al.*, 2021; H. Yang *et al.*, 2012) y la transformación logarítmica de absorbancia (J. Liu *et al.*, 2020; Qi *et al.*, 2018). Además, en las correcciones de dispersión se incluyen las derivadas espectrales, representadas por técnicas como: las derivados de Norris-Williams (Norris & Williams, 1984), el filtros de derivados polinómicos de Savitzky-Golay (SG), las derivadas gap (Slaughter *et al.*, 2001), además de las derivadas de diferencias finitas convencionales. Las transformaciones derivativas eliminan tanto efectos multiplicativos como aditivos, mejoran la resolución espectral y elimina efectos de fondo (Stevens & Ramirez-Lopez, 2013).

En estudios de suelo, se ha encontrado que la primera derivada corrige la línea base y resalta características espectrales. Mientras que la segunda derivada además corrige la tendencia lineal (Buddenbaum & Steffens, 2012; J. Liu *et al.*, 2020; Rinnan *et al.*, 2009). Asimismo, de acuerdo con Shepherd & Walsh (2002), las transformaciones derivadas minimiza el efecto de la variación en la molienda de la muestra y factores sistemáticos como variaciones de medición producto de la configuración óptica de los equipos espectrales. En el campo de la espectroscopia de suelos, no existe un acuerdo general sobre qué pretratamiento es el más efectivo. Se deben probar diferentes opciones, incluida la combinación de varias estrategias (Barra *et al.*, 2021).

2.3.4.3. Métodos de Selección de variables

Los métodos de selección de variables (MSV) son técnicas que permiten identificar y seleccionar las variables (λ) más relevantes en un conjunto de datos. Esto ayuda a mejorar el rendimiento del modelo, reducir la dimensionalidad del conjunto de datos y facilitar la interpretación de los resultados.

Los diferentes MSV ofrecen enfoques variados, el método empírico o gráfico consiste en identificar gráficamente los picos de absorción más sobresalientes y relacionarlo con su correspondiente longitud de onda, para después utilizarlos

como entradas a los modelos. En el desarrollo de prototipos de medición in situ, la selección se reduce a las λ disponibles en los elementos emisores de luz del sistema en cuestión (Masrie *et al*, 2018; Masrie *et al*, 2019; Zhou *et al.*, 2019; etc). Los métodos más sofisticados son combinaciones de técnicas de optimización, de análisis estadístico y de regresión multivariante. En el Cuadro 2.7 se muestra una revisión de los métodos utilizados en análisis de suelos con espectroscopia. Estos MVS emplean diferentes procesos, una tipificación de estos métodos en espectros es presentada por Yun, *et al*, (2019). Donde se identifican cuatro características principales en los MSV:

a) Inicialización de variables. Los MSV pueden iniciar tomando en cuenta la totalidad de las λ del espectro, y posteriormente ir descartando una a una las bandas menos informativas con respecto a una métrica de evaluación. Un ejemplo de este tipo de inicialización es el que utiliza el Método Iterativo de Eliminación Escalonada (ISE por sus siglas en inglés) que se ha utilizado en la detección de Carbono y Nitrógeno en suelos (Kawamura *et al.*, 2017). Por otro, lado existen métodos que emplean un subconjunto de bandas generado a partir de muestreos aleatorios, tales como: Monte Carlo, Muestro Bootstrap o matrices binarias. Ejemplos de este tipo de métodos utilizados en análisis de suelo son: Algoritmos Genéticos (AG) (Arcos *et al.*, 1997), Eliminación de Variables no Informativas UVE (H. Yang *et al.*, 2012), Estrategia de Eliminación Iterativa de Variables (IRIV) (Y. H. Yun *et al.*, 2014), Muestreo Adaptativo Competitivo Reponderado (CARS) (Vohland *et al.*, 2014), etc.

b) Método de calibración. Son métodos para relacionar un subconjunto de variables en el desempeño de un modelo de regresión, para cada subconjunto se calibra un modelo. Existen diversos enfoques de regresión que pueden ser utilizados en espectros de suelos. Los métodos de regresión multivariante como la regresión con Componentes Principales (PCR), Mínimos Cuadrados Parciales (PLSR) o con Operador de Selección y Contracción Mínima Absoluta (LASSO) son los más comunes. Además la Regresión de Vectores de Soporte también se ha utilizado para modelar relaciones de propiedades del suelo como contenido

de Nitrógeno (Zhang, *et al* 2016) y de Materia Orgánica (Yang, *et al*, 2019) y es posible utilizarlo como método de modelado en MSV al igual que las Redes Neuronales Artificiales aunque con un costo computacional mayor.

Cuadro 2.7 Métodos de selección de variables para espectroscopia de suelos.

Método	Elemento	Referencia
Eliminación continua análisis wavelet (CRW)-SVM	N	Zhang Y, <i>et al</i> , 2016
CARS-PLS	N	Nie, <i>et al</i> , 2017
Uninformative variable selection (UVE)-PLS	N	Nie, <i>et al</i> , 2017
ISE-PLS	CT y N	Kawamura K. <i>et al</i> 2017
Algoritmo de recorte de pico iterativo no lineal sensible a estadísticas (SNIP)	CT	Dotto A. <i>et al</i> , 2017
SNIP	Arena, limo, arcilla	Dotto A. <i>et al</i> , 2017
Algoritmo de proyecciones sucesivas (SPA)-PCR		
AG-PCR	MO	Lin Z, <i>et al</i> , 2017
CARS-PLS		
Red neuronal elástica (ENET)-PLS	MO	Hong Y. <i>et al</i> , 2018
AG -PLS		
Ga-Red de aprendizaje extremo (ELM)		
AG-PLS	MO, PH	Yang M. <i>et al</i> , 2018
AG-Modelo Cubista		
AG-SVMR		
Random Frog Leaping (RFL)- Red n. Wavelet (WNN)	N	Yao <i>et al</i> , 2018
Ant Colony Optimization (ACO) -Información mutua (MI)	N	Zhang <i>et al</i> , 2019
AG-SPA		
CARS-SPA	MO	Yuan J. <i>et al</i> , 2020
Optimización de combinación de intervalos (ICO)-SPA		
Análisis de sensibilidad (SA)-SPA		
CARS-PCR		
CARS-PLS	CO	Chen H. <i>et al</i> , 2020
CARS-BPN-DL		
SPA-PLS	P, K	Guo <i>et al.</i> , 2021

MO=materia orgánica, CT= Carbono Total, CO = Carbono orgánico.

c) Métrica de evaluación. El rendimiento de predicción de un subconjunto de variables en el modelo se calcula en función de una métrica de evaluación. En análisis de suelos, se emplean las métricas descritas en el Cuadro 2.4.

d) Estrategia de selección. Esta etapa tiene como objetivo encontrar el subconjunto de características óptimo buscando en el espacio de variables. Los métodos utilizados para seleccionar los subconjuntos se basan en:

- **Filtros.** Se realiza una evaluación de estabilidad de las variables y se rankean en orden ascendente, después establece un umbral de corte basado en alguna métrica de ajuste del modelo, como el comportamiento del RMSE.
- **Valores extremos;** Sólo se seleccionan los modelos cuyas variables contribuyan a un RMSE mínimo o a un alto coeficiente de regresión.
- **Búsqueda secuencial.** Es la estrategia usada en selección de variables hacia adelante y/o hacia atrás. Para la selección hacia adelante, las variables se agregan secuencialmente a un conjunto de variables candidatas en un modelo hasta que la adición de más variables no disminuye el criterio de evaluación (como el error). Para la selección hacia atrás, las variables se eliminan secuencialmente de un conjunto completo, hasta que la eliminación de variables no aumenta la métrica.
- **Algoritmos de optimización inteligente;** Mediante submuestreo estadístico o validación cruzada, estos métodos evalúan múltiples modelos, utilizando algoritmos evolutivos y de enjambre. Esta búsqueda utiliza una función objetivo para evaluar múltiples subconjuntos de variables de acuerdo con su rendimiento predictivo. Los algoritmos evolutivos como los GA y de enjambre como ACO son ejemplos típicos.
- **Análisis de población:** Esta búsqueda extrae información estadística de una población de submodelos que se construyen con una población de subconjuntos de variables. Realiza un análisis estadístico con los resultados de varios parámetros generados. Considera la salida de

interés como una distribución mediante la cual se pueden realizar varias pruebas de significación estadística paramétricas y/o no paramétricas.

2.3.4.4. Modelado

En esta etapa se busca a través de métodos de regresión multivariante, relacionar la absorción de las variables espectrales con las propiedades de interés en el suelo, ya sea con el espectro completo o con variables seleccionadas. Esto implica establecer una relación entre dos matrices; la primera formada por las variables predictoras X (longitudes de onda) y la segunda por las variables a predecir Y (NPK).

En técnicas de aprendizaje automático para espectroscopia de suelos, existen diferentes enfoques y técnicas para modelar y comprender la relación entre múltiples variables predictoras y la variable objetivo. Se utiliza generalmente aprendizaje supervisado para la calibración de estos modelos. Los modelos de regresión para estimar los nutrientes en el suelo comprenden algoritmos que presuponen tanto relaciones lineales como no lineales, en el Cuadro 2.8 se presentan los métodos de calibración utilizados para relacionar diversos elementos del suelo, así como los preprocesamientos espectrales, la técnica espectral, y la selección de variables involucradas para cada caso.

Aunque los métodos de calibración multivariante de relación lineal, como el análisis de componentes principales (PCA), la regresión por mínimos cuadrados parciales (PLS), son técnicas muy utilizadas en ciencias del suelo, la elección del algoritmo de regresión para predecir los niveles de nutrientes NPK en el suelo es un proceso crítico en la calibración de modelos espectrales.

Dado que los niveles de NPK pueden estar influenciados por una serie de factores complejos y no lineales, se hace necesario evaluar cuidadosamente la viabilidad de los métodos de regresión disponibles. En este contexto, los modelos de regresión no lineal, como las redes neuronales artificiales (ANN), las máquinas de soporte vectorial (SVM), o los árboles de decisión descritos en la sección 2.2.3, pueden resultar particularmente beneficiosos, ya que permiten capturar relaciones altamente no lineales entre las características espectrales y los niveles de nutrientes.

Cuadro 2.8. Métodos de regresión para detección de NPK en suelos.

Elemento	Técnica	Pretratamiento	MSV	Regresión	Métricas	Referencia
N, P, K	MIR	Derivativo, SGs	-----	PLS	RMSE, R ² , RPD, PR	(Ji <i>et al.</i> , 2016)
N	Vis-NIR	Normalización, SGs, SNV	-----	PLS, PCR, LS-SVM, CRM	RMSE, R ² , RPD	(Morellos <i>et al.</i> , 2016)
N	NIR	WDD	CR	SVM	RMSE, R ² , RPD	(Zhang <i>et al.</i> , 2016)
N	Vis-NIR	EC	SGd1	PLS	RMSE, R ² , RPD	(Tümsavaş, 2017)
N	NIR	----	UVE, CARS	PLS, LS-SVM	R ² , RMSE	(Nie <i>et al.</i> , 2017)
N, P, K	Vis-NIR	SGs, SGd1, SGd2, WT, SNV, MSC, LG	----	BPNN, PLS	SSR/SST, RPD	(Qi <i>et al.</i> , 2018)
N	NIR	SGs	RFL	WNN	RMSE, R ²	(Yao <i>et al.</i> , 2018)
N	Vis-NIR	SASI	C. de correlación	BPNN	RMSE, R ² , RPD	(Song <i>et al.</i> , 2019)
N	NIR	SGs, SGd1, DT, SNV	AG, SPA, CARS, BI	PLS	RMSE, R ² , RPD	(Xiao & He, 2019)
N, P, K	Vis-NIR	MSC, SGd1, SGd2	C. de correlación	PLS	RMSE, R ² , RPIQ	(Coutinho <i>et al.</i> , 2019)
P, K	XRF	-----	RF	SGLM	RMSE, R ² , RPD	(Andrade <i>et al.</i> , 2020)
N, P, K	NIR	-----	-----	BPNN	R ²	(MacAbiog <i>et al.</i> , 2020)
N, P, K	Vis-NIR	SGs	DASU	MLR	R ²	(Patel, <i>et al.</i> , 2020)
N	Vis-NIR	SGs, SGd1, SGd2, SNV, MSC	AG	PCR	RMSE, R ² , RPD	(Yubing Wang <i>et al.</i> , 2021)
P, K	Vis-NIR	MAS, SGs, SGd1, SNV, normalización	-----	PLS	RMSE, R ² , RPD	(Munnaf <i>et al.</i> , 2021)
N	MIR	MSC, SGs, SGd1, SGd2, SNV	-----	SVM	R ² , RPIQ	(Wehrle <i>et al.</i> , 2021)

P, K	Vis-NIR	SGs, MSC, FD-LGR	SPA, CARS	PLS, SVM	RMSE, R ²	(Guo <i>et al.</i> , 2021)
N	MIR	SNV, MSC	CARS, BOSS	PLS, MLR	RMSE, R ² , RPD	(Hong Li <i>et al.</i> , 2022)
P	NIR	SNV	-----	PLS	R ² , RPIQ, RPD	(Cao <i>et al.</i> , 2022)
P, K	Vis-NIR	SGs, SNV, MSC, DT	VIP, PCA	PLS	RMSE, R ² , RPD	(Mammado <i>v et al.</i> , 2022)
N	Vis-NIR	SGs, MSC, SD, WDD	PCA	PLS, BPNN, ELM, SVM, CNN	RMSE, R ²	(J. Ma <i>et al.</i> , 2022)
N, P, K	Vis-NIR	SGs	SNR		RMSE, R ²	(Jiang <i>et al.</i> , 2023)

PLS=Regresión por Mínimos Cuadrados Parciales, CRM=Modelo de Regresión Cubista, SVM=Máquina de Soporte de Vectores, LS-SVM=Mínimos Cuadrados con Máquina de Soporte de Vectores, PCA=Análisis de Componentes Principales, CARS= Algoritmo Competitivo Reponderado Adaptativo, BPNN=Red Neuronal de Retropropagación, WNN=Red Neuronal Wavelet, SGLM= Modelos Lineales Generalizados por Pasos, RF= Bosques Aleatorios, PR = reproducibilidad de predicción, LG = transformación logarítmica, WT= transformación wavelet, SASI= Índice espectral ajustado del suelo; BI = intervalo hacia atrás, XRF = Fluorescencia de rayos X, DASU = Método de análisis derivado para desmezcla espectral, MAS = suavizado de promedio móvil, FD-LGR = derivadas de primer orden basadas en transformaciones logarítmicas e inversas, BOSS = enfoque de contracción suave de arranque, VIP = Importancia de la variable en la proyección, SD = derivadas espectrales, MLR= regresión lineal múltiple, WDD = Eliminación de ruido de dominio wavelet, CNN = redes neuronales convolucionales, SNR = relación señal – ruido, ELM = red de aprendizaje extremo.

La incorporación de información contextual y la utilización de características espectrales relevantes en un modelo no lineal podrían mejorar significativamente la precisión y capacidad de generalización de las predicciones. Sin embargo, la elección del modelo adecuado debe basarse en una comparación exhaustiva de diferentes enfoques y una validación rigurosa utilizando conjuntos de datos independientes para garantizar que el modelo final sea capaz de proporcionar estimaciones confiables de los niveles de NPK en diferentes condiciones de suelo y ubicaciones geográficas.

Referencias del capítulo

- Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1). <https://doi.org/10.1016/j.jjimei.2020.100004>
- Ahmadi, A., Emami, M., Daccache, A., & He, L. (2021). Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis. *Agronomy*, 11(3). <https://doi.org/10.3390/agronomy11030433>
- Albawi, S., & Mohammed, T. A. (2017). Understanding of a Convolutional Neural Network. *IEEE*. <https://ieeexplore.ieee.org/abstract/document/8308186>
- Algarni, Z. Y. (2019). Variable selection in count data regression model based on firefly algorithm. *Statistics, Optimization and Information Computing*, 7(2), 520–529. <https://doi.org/10.19139/soic.v7i2.566>
- Alibabaei, K., Gaspar, P. D., Lima, T. M., Campos, R. M., Girão, I., Monteiro, J., & Lopes, C. M. (2022). A Review of the Challenges of Using Deep Learning Algorithms to Support Decision-Making in Agricultural Activities. *Remote Sensing*, 14(3), 1–43. <https://doi.org/10.3390/rs14030638>
- Andrade, R., Faria, W. M., Silva, S. H. G., Chakraborty, S., Weindorf, D. C., Mesquita, L. F., Guilherme, L. R. G., & Curi, N. (2020). Prediction of soil fertility via portable X-ray fluorescence (pXRF) spectrometry and soil texture in the Brazilian Coastal Plains. *Geoderma*, 357(July 2019), 113960. <https://doi.org/10.1016/j.geoderma.2019.113960>
- Araújo, S. R., Wetterlind, J., Demattê, J. A. M., & Stenberg, B. (2014). Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. *European Journal of Soil Science*, 65(5), 718–729. <https://doi.org/10.1111/ejss.12165>
- Arcos, J. M., Cruz Ortiz, M., Villahoz, B., & Sarabia, L. A. (1997). Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: Application on copper and zinc mixture. *Talanta*, 59(2), 311–317. [https://doi.org/10.1016/S0039-9140\(02\)00505-2](https://doi.org/10.1016/S0039-9140(02)00505-2)
- Armenise, E., Redmile-Gordon, M. A., Stellacci, A. M., Ciccarese, A., & Rubino, P. (2013). Developing a soil quality index to compare soil fitness for agricultural use under different managements in the mediterranean environment. *Soil and Tillage Research*, 130, 91–98. <https://doi.org/10.1016/j.still.2013.02.013>
- Atasoy, S., Mateus, D., Georgiou, A., Navab, N., & Yang, G. Z. (2011). Wave interference for pattern description. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6493 LNCS(PART 2), 41–54. https://doi.org/10.1007/978-3-642-19309-5_4
- Baek, J. Y., Yoo, Y. S., & Bae, S. H. (2019). Adversarial Learning With Knowledge of Image Classification for Improving GANs. *IEEE Access*, 7, 56591–56605. <https://doi.org/10.1109/ACCESS.2019.2913697>
- Barber, D. (2020). *Bayesian Reasoning and Machine Learning*.
- Barefoot, A., Murphy, J., & Aizawa, H. (2003). *Handbook of Residue Analytical Methods for Agrochemicals* (P. W. Lee (ed.); Wiley).
- 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>

- Bartholdi, E., & Ernst, R. R. (1973). Fourier spectroscopy and the causality principle. *Journal of Magnetic Resonance* (1969), 11(1), 9–19. [https://doi.org/10.1016/0022-2364\(73\)90076-0](https://doi.org/10.1016/0022-2364(73)90076-0)
- Bellon-Maurel, V., Fernandez-ahumada, E., Palagos, B., Roger, J., & Mcbratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry*, 29(9), 1073–1081. <https://doi.org/10.1016/j.trac.2010.05.006>
- Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., & Sommer, S. (2009). Using Imaging Spectroscopy to study soil properties. *Remote Sensing of Environment*, 113(SUPPL. 1), S38–S55. <https://doi.org/10.1016/j.rse.2008.09.019>
- Ben-dor, E., & Epema, G. F. (2014). *Soil Reflectance. January 1999.*
- Ben Dor, E., Ong, C., & Lau, I. C. (2015). Reflectance measurements of soils in the laboratory: Standards and protocols. *Geoderma*, 245–246, 112–124. <https://doi.org/10.1016/j.geoderma.2015.01.002>
- Brame, C. J. (2007). A guide To active learning -Vanderbilt white paper. *Vanderbilt University Center for Teaching*. <https://cft.vanderbilt.edu/active-learning/>
- Brieman, L. (1996). Bagging predictors. *Machine Learning*, 24(3), 123–140. <https://doi.org/10.3390/risks8030083>
- Bronick, C. J., & Lal, R. (2021). Soil structure and management : a review. *Geoderma*, 124(2005), 3–22. <https://doi.org/10.1016/j.geoderma.2004.03.005>
- Brunton, S. L., & Kutz, J. N. (2019). *Data-Driven Science and Engineering* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108380690>
- Buchröller, W., & Tye, S. H. H. (1980). Vibrational states in the spectroscopy. *Physical Review Letters*, 44(13), 850–853. <https://doi.org/10.1103/PhysRevLett.44.850>
- Buddenbaum, H., & Steffens, M. (2012). *The Effects of Spectral Pretreatments on Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy. 2012.* <https://doi.org/10.1155/2012/274903>
- Burger, J., & Geladi, P. (2007). Spectral pre-treatments of hyperspectral near infrared images: Analysis of diffuse reflectance scattering. *Journal of Near Infrared Spectroscopy*, 15(1), 29–37. <https://doi.org/10.1255/jnirs.717>
- Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J., & Smaling, E. M. A. (2012). Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. *Geoderma*, 183–184, 41–48. <https://doi.org/10.1016/j.geoderma.2012.03.011>
- Cameron, E. K., Martins, I. S., Lavelle, P., Mathieu, J., Tedersoo, L., Gottschall, F., Guerra, C. A., Hines, J., Patoine, G., Siebert, J., Winter, M., Cesarz, S., Delgado-Baquerizo, M., Ferlian, O., Fierer, N., Kreft, H., Lovejoy, T. E., Montanarella, L., Orgiazzi, A., ... Eisenhauer, N. (2018). Global gaps in soil biodiversity data. *Nature Ecology and Evolution*, 2(7), 1042–1043. <https://doi.org/10.1038/s41559-018-0573-8>
- Cao, Z., Kühn, P., He, J.-S., Bauhus, J., Guan, Z.-H., & Scholten, T. (2022). Calibration of Near-Infrared Spectra for Phosphorus Fractions in Grassland Soils on the Tibetan Plateau. *Agronomy*, 12(4), 783. <https://doi.org/10.3390/agronomy12040783>
- Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1), 103–112. <https://doi.org/10.1016/j.chemolab.2004.12.011>
- Ciani, A., Goss, K. U., & Schwarzenbach, R. P. (2005). Light penetration in soil and particulate minerals. *European Journal of Soil Science*, 56(5), 561–574. <https://doi.org/10.1111/j.1365-2389.2005.00688.x>

- Clark, R. N., King, T. V. V., Klejwa, M., Swayze, G. A., & Vergo, N. (1990). High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research*, 95(B8). <https://doi.org/10.1029/jb095ib08p12653>
- Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, 89(B7), 6329–6340. <https://doi.org/10.1029/JB089iB07p06329>
- Cootes, T. F., Ionita, M. C., Lindner, C., & Sauer, P. (2012). Robust and accurate shape model fitting using random forest regression voting. In A. Fitzgibbon (Ed.), *Lecture Notes in Computer Science: Vol. 7578 LNCS* (Issue PART 7, pp. 278–291). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33786-4_21
- Coutinho, M. A. N., Alari, F. de O., Ferreira, M. M. C., & Amaral, L. R. d. (2019). Influence of soil sample preparation on the quantification of NPK content via spectroscopy. *Geoderma*, 338(July 2018), 401–409. <https://doi.org/10.1016/j.geoderma.2018.12.021>
- D. C. Slaughter, M. G. Pelletier, & S. K. Upadhyaya. (2001). Sensing Soil Moisture Using Nir Spectroscopy. *Applied Engineering in Agriculture*, 17(2), 241–247. <https://doi.org/10.13031/2013.5449>
- D'Acqui, L. P., Pucci, A., & Janik, L. J. (2010). Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. *European Journal of Soil Science*, 61(6), 865–876. <https://doi.org/10.1111/j.1365-2389.2010.01301.x>
- Dhaliwal, S. S., Naresh, R. K., Mandal, A., Walia, M. K., Gupta, R. K., Singh, R., & Dhaliwal, M. K. (2019). Effect of manures and fertilizers on soil physical properties, build-up of macro and micronutrients and uptake in soil under different cropping systems: a review. *Journal of Plant Nutrition*, 42(20), 2873–2900. <https://doi.org/10.1080/01904167.2019.1659337>
- Ding, J., Yang, A., Wang, J., Sagan, V., & Yu, D. (2018). Machine-learning-based quantitative estimation of soil organic carbon content by VIS/NIR spectroscopy. *PeerJ*, 2018(10), 1–24. <https://doi.org/10.7717/peerj.5714>
- Domingos, P. (2015). *The Master Algorithm*. Perseus Books Group.
- Donald, P., Gary, L., Kriz, G., & Vyvyan, J. (2017). Introduction to Spectroscopy. In *Spectroscopic Methods in Food Analysis* (5th ed.). Cengage Learning. <https://doi.org/10.1201/b21879-2>
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241–258. <https://doi.org/10.1007/s11704-019-8208-z>
- Dotto, A. C., Dalmolin, R. S. D., Grunwald, S., ten Caten, A., & Pereira Filho, W. (2017). Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. *Soil and Tillage Research*, 172(December 2015), 59–68. <https://doi.org/10.1016/j.still.2017.05.008>
- Ehrenfeld, J. G., Ravit, B., & Elgersma, K. (2005). Feedback in the plant-soil system. *Annual Review of Environment and Resources*, 30, 75–115. <https://doi.org/10.1146/annurev.energy.30.050504.144212>
- Eiben, A. E. (2015). Introduction to Evolutionary Computing. In *Kybernetes* (Vol. 33). <https://doi.org/10.1108/03684920410699216>
- El-Zeiny, M. B., Zawbaa, H. M., & Serag, A. (2021). An evaluation of different bio-inspired feature selection techniques on multivariate calibration models in spectroscopy. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 246, 119042. <https://doi.org/10.1016/j.saa.2020.119042>
- Fei, Q., Li, M., Wang, B., Huan, Y., Feng, G., & Ren, Y. (2009). Analysis of cefalexin with NIR spectrometry coupled to artificial neural networks with modified genetic

- algorithm for wavelength selection. *Chemometrics and Intelligent Laboratory Systems*, 97(2), 127–131. <https://doi.org/10.1016/j.chemolab.2009.03.003>
- Gholizadeh, A., Saberioon, M., Ben-Dor, E., Viscarra Rossel, R. A., & Borůvka, L. (2020). Modelling potentially toxic elements in forest soils with vis–NIR spectra and learning algorithms. *Environmental Pollution*, 267. <https://doi.org/10.1016/j.envpol.2020.115574>
- Gobrecht, A., Roger, J. M., & Bellon-Maurel, V. (2014). Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation. A Review. In *Advances in Agronomy* (Vol. 123). <https://doi.org/10.1016/B978-0-12-420225-2.00004-2>
- Gondal, A. H., Hussain, I., Ijaz, A. bakar, Zafar, A., Ch, B. I., Zafar, H., Sohail, M. D., Niazi, H., Touseed, M., Khan, A. A., Tariq, M., Yousuf, H., & Usama, M. (2021). Influence of soil pH and microbes on mineral solubility and plant nutrition: A review . *International Journal of Agriculture and Biological Sciences*, 5(1), 2–12. https://d1wqtxts1xzle7.cloudfront.net/66113631/Volume_5_Issue_1_Paper_8-with-cover-page-v2.pdf?Expires=1639806217&Signature=FRqt9IzQuEtA1LG0eYffRC1830B80os5kz3-tDqJEqNJJWAdYnhZ8L00~uKz-5hrWLUZmCUQgimABd9LjVhOc1X44eq-W-rC8HtbTBEQeJWWTwNzo0D97Di4NnH1DQcBNJY
- Gonzalez, A., & Vargas, N. (2014). Análisis de imágenes hiperespectrales To cite this version : HAL Id : hal-00935014. *Hal*, 9, 14–17. <https://hal.archives-ouvertes.fr/hal-00935014/document>
- González, M., & Montaña, L. (2015). La espectroscopia y su tecnología: Un repaso histórico y su importancia para el siglo XXI. *Latin-American Journal of Physics Education*, 9(4), 13.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144. <https://doi.org/10.1145/3422622>
- Guiñón, J. L., Ortega, E., García-Antón, J., & Pérez-herranz, V. (2007). Moving Average and Savitzki-Golay Smoothing Filters Using Mathcad. *International Conference on Engineering Education*, 1, 1–4. <http://academic.research.microsoft.com/Paper/12119855.aspx>
- Guo, P., Li, T., Gao, H., Chen, X., Cui, Y., & Huang, Y. (2021). Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using vis-nir spectroscopy. *Remote Sensing*, 13(19), 1–20. <https://doi.org/10.3390/rs13194000>
- Hastie, T., Tibshirani, R., & Friedman, J. (2005). The Elements of Statistical Learning : Data Mining , Inference and Prediction Probability Theory : The Logic of Science The Fundamentals of Risk Measurement Mathematicians , pure and applied , think there is something weirdly different about. *The Mathematical Intelligencer*, 27(2), 83–85. <http://link.springer.com/article/10.1007/BF02985802?LI=true#>
- He, Y., Huang, M., García, A., Hernández, A., & Song, H. (2007). Prediction of soil macronutrients content using near-infrared spectroscopy. *Computers and Electronics in Agriculture*, 58(2), 144–153. <https://doi.org/10.1016/j.compag.2007.03.011>
- He, Y., Xiao, S., Nie, P., Dong, T., Qu, F., & Lin, L. (2017). Research on the Optimum Water Content of Detecting Soil Nitrogen Using Near Infrared Sensor. *Sensors (Switzerland)*, 17(2045), 1–12. <https://doi.org/10.3390/s17092045>
- Helland, I. S. (1990). Partial Least Squares Regression and Statistical Models. *Scandinavian Journal of Statistics*, 17(2), 97–114. <http://www.jstor.org/stable/4616159>

- Hollas, J. M. (2011). Modern Spectroscopy. In *Voenno-meditsinskiĭ zhurnal* (Vol. 332, Issue 2).
- Hong, Y., Chen, Y., Yu, L., Liu, Y., Liu, Y., Zhang, Y., Liu, Y., & Cheng, H. (2018). Combining fractional order derivative and spectral variable selection for organic matter estimation of homogeneous soil samples by VIS-NIR spectroscopy. *Remote Sensing*, 10(3). <https://doi.org/10.3390/rs10030479>
- Horta, M. D. C., & Torrent, J. (2007). The Olsen P method as an agronomic and environmental test for predicting phosphate release from acid soils. *Nutrient Cycling in Agroecosystems*, 77(3), 283–292. <https://doi.org/10.1007/s10705-006-9066-2>
- Huang, G., Zhu, Q., & Siew, C. (2006). Extreme learning machine : Theory and applications. *Neurocomputing*, 70, 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
- Inoue, Y. (2020a). Satellite- and drone-based remote sensing of crops and soils for smart farming—a review. *Soil Science and Plant Nutrition*, 66(6), 798–810. <https://doi.org/10.1080/00380768.2020.1738899>
- Inoue, Y. (2020b). Soil Science and Plant Nutrition Satellite- and drone-based remote sensing of crops and soils for smart farming – a review. *Soil Science and Plant Nutrition*, 66(6), 798–810. <https://doi.org/10.1080/00380768.2020.1738899>
- Isaac, W., & Na, A. (2017). On-the-go soil nitrogen sensor based on near infrared spectroscopy. *2016 International Conference on Information Technology, IncITe 2016 - The Next Generation IT Summit on the Theme - Internet of Things: Connect Your Worlds*, 312–315. <https://doi.org/10.1109/INCITE.2016.7857637>
- Isaak, S., Yusof, Y., Ngajikin, N. H., Ramli, N., & Wen, C. M. (2019). A low cost spectroscopy with Raspberry Pi for soil macronutrient monitoring. *Telkomnika (Telecommunication Computing Electronics and Control)*, 17(4), 1867–1875. <https://doi.org/10.12928/TELKOMNIKA.V17I4.12775>
- Ji-yong, S., Xiao-bo, Z., Xiao-wei, H., Jie-wen, Z., Yanxiao, L., Limin, H., & Jianchun, Z. (2013). Rapid detecting total acid content and classifying different types of vinegar based on near infrared spectroscopy and least-squares support vector machine. 138, 192–199. <https://doi.org/10.1016/j.foodchem.2012.10.060>
- Ji, W., Adamchuk, V. I., Biswas, A., Dhawale, N. M., Sudarsan, B., Zhang, Y., Viscarra Rossel, R. A., & Shi, Z. (2016). Assessment of soil properties in situ using a prototype portable MIR spectrometer in two agricultural fields. *Biosystems Engineering*, 152, 14–27. <https://doi.org/10.1016/j.biosystemseng.2016.06.005>
- Ji, W., Li, S., Chen, S., Shi, Z., Viscarra Rossel, R. A., & Mouazen, A. M. (2016). Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. *Soil and Tillage Research*, 155, 492–500. <https://doi.org/10.1016/j.still.2015.06.004>
- Jiang, C., Zhao, J., Ding, Y., & Li, G. (2023). Vis-NIR Spectroscopy Combined with GAN Data Augmentation for Predicting Soil Nutrients in Degraded Alpine Meadows on the Qinghai-Tibet Plateau. *Sensors (Basel, Switzerland)*, 23(7). <https://doi.org/10.3390/s23073686>
- Jin, J., Chen, Z., Li, L., Steponavicius, R., Thennadil, S. N., Yang, J., & Yu, R. (2012). Quantitative Spectroscopic Analysis of Heterogeneous Mixtures : The Properties of Samples. *Analytical Chemistry*, 320–326.
- Kang, H., Gao, H., & Yu, W. (2017). Evaluation of Spectral Pretreatments , Spectral Range and Regression Methods for Quantitative Spectroscopic Analysis of Soil Organic Carbon Composition. *Spectroscopy Letters*, 7010(March). <https://doi.org/10.1080/00387010.2017.1297956>
- Kawamura, K., Nishigaki, T., Tsujimoto, Y., Andriamananjara, A., Rabenaribo, M., Asai,

- H., Rakotoson, T., & Razafimbelo, T. (2021). Exploring relevant wavelength regions for estimating soil total carbon contents of rice fields in Madagascar from Vis-NIR spectra with sequential application of backward interval PLS. *Plant Production Science*, 24(1), 1–14. <https://doi.org/10.1080/1343943X.2020.1785898>
- Kawamura, K., Tsujimoto, Y., Rabenarivo, M., Asai, H., Andriamananjara, A., & Rakotoson, T. (2017). Vis-NIR spectroscopy and PLS regression with waveband selection for estimating the total C and N of paddy soils in Madagascar. *Remote Sensing*, 9(10). <https://doi.org/10.3390/rs9101081>
- Keyvan, K., Sohrabi, M. R., & Motiee, F. (2021). An intelligent method based on feed-forward artificial neural network and least square support vector machine for the simultaneous spectrophotometric estimation of anti hepatitis C virus drugs in pharmaceutical formulation and biological fluid. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 263, 120190. <https://doi.org/10.1016/j.saa.2021.120190>
- Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., & Abbas, A. (2018). Modern Trends in Hyperspectral Image Analysis: A Review. *IEEE Access*, 6(c), 14118–14129. <https://doi.org/10.1109/ACCESS.2018.2812999>
- Kim, H. J., Sudduth, K. A., & Hummel, J. W. (2009). Soil macronutrient sensing for precision agriculture. *Journal of Environmental Monitoring*, 11(10), 1810–1824. <https://doi.org/10.1039/b906634a>
- Kose, U., Prasath, V. B. S., Mondal, M. R. H., Podder, P., & Bharati, S. (2022). Artificial Intelligence and Smart Agriculture Applications. In *Artificial Intelligence and Smart Agriculture Applications*. <https://doi.org/10.1201/9781003311782>
- Laskar, S., & Mukherjee, S. (2016). Optical Sensing Methods for Assessment of Soil Macro-nutrients and other Properties for Application in Precision Agriculture: A review. *ADBU-Journal of Engineering Technology AJET*, 4(1), 206.
- Legnaioli, S., Lorenzetti, G., Pardini, L., Cavalcanti, G. H., & Palleschi, V. (2014). Double and multiple pulse LIBS techniques. In *Laser-Induced Breakdown Spectroscopy* (Vol. 182, pp. 117–141). https://doi.org/10.1007/978-3-642-45085-3_5
- Lei, T., & Sun, D. W. (2022). Achieving joint calibration of soil Vis-NIR spectra across instruments, soil types and properties by an attention-based spectra encoding-spectra/property decoding architecture. *Geoderma*, 405(May 2021), 115449. <https://doi.org/10.1016/j.geoderma.2021.115449>
- Lei, Z., Yao, M., Liu, M., Li, Q., & Mao, H. (2011). Comparison between fertilization N, P, K and No fertilization N, P, K in paddy soil by laser induced breakdown spectroscopy. *Proceedings - 4th International Conference on Intelligent Computation Technology and Automation, ICICTA 2011*, 1, 363–366. <https://doi.org/10.1109/ICICTA.2011.102>
- Li, H., Jia, S., & Le, Z. (2019). Quantitative Analysis of Soil Total Nitrogen Using Hyperspectral Imaging Technology with Extreme Learning Machine. *Sensors (Switzerland)*.
- Li, Hong, Wang, J., Zhang, J., Liu, T., Acquah, G. E., & Yuan, H. (2022). Combining Variable Selection and Multiple Linear Regression for Soil Organic Matter and Total Nitrogen Estimation by DRIFT-MIR Spectroscopy. *Agronomy*, 12(3). <https://doi.org/10.3390/agronomy12030638>
- Li, Y., Yang, Q., Chen, M., Wang, M., & Zhang, M. (2019). An ISE-based On-Site Soil Nitrate Nitrogen Detection System. *Sensors (Switzerland)*, 3.
- Liakos, K. G., Busato, P., Moshou, D., & Pearson, S. (2018). Machine Learning in Agriculture : A Review. *Sensors (Basel, Switzerland)*, 18, 1–29. <https://doi.org/10.3390/s18082674>

- Lin, Z. D., Wang, Y. B., Wang, R. J., Wang, L. S., Lu, C. P., Zhang, Z. Y., Song, L. T., & Liu, Y. (2017). Improvements of the Vis-NIRS Model in the Prediction of Soil Organic Matter Content Using Spectral Pretreatments, Sample Selection, and Wavelength Optimization. *Journal of Applied Spectroscopy*, 84(3), 529–534. <https://doi.org/10.1007/s10812-017-0505-4>
- Linker, R., Shmulevich, I., Kenny, A., & Shaviv, A. (2005). Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy. *Chemosphere*, 61(5), 652–658. <https://doi.org/10.1016/j.chemosphere.2005.03.034>
- Liu, J., Xie, J., Han, J., Wang, H., Sun, J., Li, R., & Li, S. (2020). Visible and near-infrared spectroscopy with chemometrics are able to predict soil physical and chemical properties. *Journal of Soils and Sediments*, 20(7), 2749–2760. <https://doi.org/10.1007/s11368-020-02623-1>
- Liu, Lanfa, Ji, M., & Buchroithner, M. (2018). Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. *Sensors (Switzerland)*, 18(9). <https://doi.org/10.3390/s18093169>
- Liu, Lijuan, Shen, B., & Wang, X. (2014). Research on kernel function of support vector machine. *Lecture Notes in Electrical Engineering*, 260 LNEE, 827–834. https://doi.org/10.1007/978-94-007-7262-5_93
- Lourembam, D., Laskar, S., & Mukherjee, S. (2017). Framework for an optical sensor system for monitoring of soil nitrogen and tailoring soil pH. *Journal of Optics*. <https://doi.org/10.1007/s12596-017-0434-x>
- Lv, N., Ma, H., Member, S., Chen, C., & Member, S. (2021). Remote Sensing Data Augmentation Through Adversarial Training. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 9318–9333. <https://doi.org/10.1109/JSTARS.2021.3110842>
- Lyon, L. A., Keating, C. D., Fox, A. P., Baker, B. E., He, L., Nicewarner, S. R., Mulvaney, S. P., & Natan, M. J. (1998). Raman Spectroscopy. *Analytical Chemistry*, 70(12), 341–361.
- Ma, F., Du, C. W., Zhou, J. M., & Shen, Y. Z. (2019). Investigation of soil properties using different techniques of mid-infrared spectroscopy. *European Journal of Soil Science*, 70(1), 96–106. <https://doi.org/10.1111/ejss.12741>
- Ma, J., Cheng, J., Wang, J., Pan, R., He, F., Yan, L., & Xiao, J. (2022). Rapid detection of total nitrogen content in soil based on hyperspectral technology. *Information Processing in Agriculture*, 9(4), 566–574. <https://doi.org/10.1016/j.inpa.2021.06.005>
- MacAbiog, R. E. N., Fadchar, N. A., & Cruz, J. C. D. (2020). Soil NPK Levels Characterization Using Near Infrared and Artificial Neural Network. *Proceedings - 2020 16th IEEE International Colloquium on Signal Processing and Its Applications, CSPA 2020, Cspa*, 141–145. <https://doi.org/10.1109/CSPA48992.2020.9068717>
- Mammadov, E., Denk, M., Riedel, F., Ka, C., Lewinska, K., Łukowiak, R., Grzebisz, W., & Mamedov, A. I. (2022). Determination of Mehlich 3 Extractable Elements with Visible and Near Infrared Spectroscopy in a Mountainous Agricultural Land , the Caucasus Mountains.
- Manolakis, D., Lockwood, R., & Cooley, T. (2016). Hyperspectral Imaging Remote Sensing_Physics, Sensors, and Algorithms. In *Hyperspectral Imaging Remote Sensing*.
- Masrie, M., Rosli, A. Z. M., Sam, R., Janin, Z., & Nordin, M. K. (2019). Integrated optical sensor for NPK Nutrient of Soil detection. *2018 IEEE 5th International Conference*

- on Smart Instrumentation, Measurement and Application, ICSIMA 2018, November, 28–30. <https://doi.org/10.1109/ICSIMA.2018.8688794>
- Masrie, M., Rosman, M. S. A., Sam, R., & Janin, Z. (2018). Detection of nitrogen, phosphorus, and potassium (NPK) nutrients of soil using optical transducer. 2017 *IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2017, 2017-Novem*(November), 1–4. <https://doi.org/10.1109/ICSIMA.2017.8312001>
- Mendes, W. de S., Sommer, M., Koszinski, S., & Wehrhan, M. (2021). Local Peatlands Spectral Data Influence in Global Spectral Modelling of Soil Organic Carbon and Total Nitrogen Using Near-Infrared Spectrum. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3983929>
- Min, M., & Lee, W. S. (2005). Determination of Significant Wavelengths and Prediction of Nitrogen Content for Citrus. *American Society of Agricultural Engineers ISSN*, 48(1998), 455–461.
- Mirzaee, S., Ghorbani-Dashtaki, S., Mohammadi, J., Asadi, H., & Asadzadeh, F. (2016). Spatial variability of soil organic matter using remote sensing data. *Catena*, 145, 118–127. <https://doi.org/10.1016/j.catena.2016.05.023>
- Mohamed, E. S., El Baroudy, A. A., El-beshbeshy, T., Emam, M., Belal, A. A., Elfadaly, A., Aldosari, A. A., Ali, A. M., & Lasaponara, R. (2020). Vis-nir spectroscopy and satellite landsat-8 oli data to map soil nutrients in arid conditions: A case study of the northwest coast of egypt. *Remote Sensing*, 12(22), 1–20. <https://doi.org/10.3390/rs12223716>
- Montesinos-López, O. A., Montesinos-López, A., & Corssa, J. (2022). Over fitting , Model Tuning , and Evaluation of Prediction Performance. In F. van Euwijk (Ed.), *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 109–139). Springer.
- Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Wiebensohn, J., Bill, R., & Mouazen, A. M. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering*, 152, 104–116. <https://doi.org/10.1016/j.biosystemseng.2016.04.018>
- Moros Portolés, J. (2007). *TRATAMIENTO NUMÉRICO DE LOS DATOS EN EL ANÁLISIS CUANTITATIVO POR ESPECTROMETRÍA VIBRACIONAL* (Vol. 3, Issue September). Universitat de Valencia.
- Mukherjee, S., & Laskar, S. (2019). Vis–NIR-based optical sensor system for estimation of primary nutrients in soil. *Journal of Optics (India)*, 48(1), 87–103. <https://doi.org/10.1007/s12596-019-00517-1>
- Munnaf, M. A., Guerrero, A., Nawar, S., Haesaert, G., Van Meirvenne, M., & Mouazen, A. M. (2021). A combined data mining approach for on-line prediction of key soil quality indicators by Vis-NIR spectroscopy. *Soil and Tillage Research*, 205(June 2020), 104808. <https://doi.org/10.1016/j.still.2020.104808>
- Murphy, K. P. (2012). Machine Learning A Probabilistic Perspective. In *Chance Encounters: Probability in Education*. https://doi.org/10.1007/978-94-011-3532-0_2
- Nawar, S., Delbecque, N., Declercq, Y., De Smedt, P., Finke, P., Verdoodt, A., Van Meirvenne, M., & Mouazen, A. M. (2019). Can spectral analyses improve measurement of key soil fertility parameters with X-ray fluorescence spectrometry? *Geoderma*, 350(May), 29–39. <https://doi.org/10.1016/j.geoderma.2019.05.002>
- Nie, P. C., Dong, T., He, Y., & Qu, F. (2017). Detection of soil nitrogen using near infrared sensors based on soil pretreatment and algorithms. *Sensors (Switzerland)*, 17(5), 1–13. <https://doi.org/10.3390/s17051102>
- Nie, P., Dong, T., He, Y., & Xiao, S. (2018). Research on the Effects of Drying

- Temperature on Nitrogen Detection of Different Soil Types by Near. *Sensors (Switzerland)*. <https://doi.org/10.3390/s18020391>
- Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B., & Montanarella, L. (2014). Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biology and Biochemistry*, 68, 337–347. <https://doi.org/10.1016/j.soilbio.2013.10.022>
- Norris, K. H., & Williams, P. C. (1984). Optimization of Mathematical Treatments of Raw Near-Infrared Signal in the Measurement of Protein in Hard Red Spring Wheat. *Cereal Chemistry*, 2, 158–165.
- Papousek, D., & Aliev, M. R. (1982). *Molecular Vibration-Rotation Spectra* (R. Zahradnik (ed.); 1st ed.). Elsevier.
- Patel, A. K., Ghosh, J. K., & Sayyad, S. U. (2020). Fractional abundances study of macronutrients in soil using hyperspectral remote sensing. *Geocarto International*, 0(0), 1–20. <https://doi.org/10.1080/10106049.2020.1720315>
- Petropoulos, G. P., Arvanitis, K., & Sigrimis, N. (2012). Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Expert Systems with Applications*, 39(3), 3800–3809. <https://doi.org/10.1016/j.eswa.2011.09.083>
- Qi, H., Paz-Kagan, T., Karnieli, A., Jin, X., & Li, S. (2018). Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. *Soil and Tillage Research*, 175(October 2017), 267–275. <https://doi.org/10.1016/j.still.2017.09.006>
- Reeves, J. B. (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*, 158(1–2), 3–14. <https://doi.org/10.1016/j.geoderma.2009.04.005>
- 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-Pérez, J. R., Marcelo, V., Pereira-Obaya, D., García-Fernández, M., & Sanz-Ablanedo, E. (2021). Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards. *Agronomy*, 11(10). <https://doi.org/10.3390/agronomy11101895>
- Rojas, R. (1996). The Backpropagation Algorithm. In *Neural Networks* (pp. 150–182). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-61068-4_7
- Sadgrove, E. J., Falzon, G., Miron, D., & Lamb, D. W. (2021). The segmented colour feature extreme learning machine: Applications in agricultural robotics. *Agronomy*, 11(11), 1–16. <https://doi.org/10.3390/agronomy11112290>
- Sáez-Plaza, P., García, A., & Martín, J. (2019). An annotation on the Kjeldahl method. *Anales de La Real Academia Nacional de Farmacia*, 85(1), 14–19.
- Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2018). Recent Advances in Recurrent Neural Networks. *Neural and Evolutionary Computing*, 1–21.
- Sarathjith, M. C., Das, B. S., Wani, S. P., & Sahrawat, K. L. (2016). Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma*, 267, 1–9. <https://doi.org/10.1016/j.geoderma.2015.12.031>
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627–1639.
- Norma Oficial Mexicana NOM-021-RECNAT-2000, Diario Oficial de la Federación (2002). http://dof.gob.mx/nota_detalle.php?codigo=717582&fecha=31/12/2002

- Shao, Y., & He, Y. (2011). Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy. *Soil Research*, 49(2), 166–172.
<https://doi.org/10.1071/SR10098>
- Shepherd, K. D., & Walsh, M. G. (2002). Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Science Society of America Journal*, 66(3), 988–998. <https://doi.org/10.2136/sssaj2002.9880>
- Shi, P., Castaldi, F., van Wesemael, B., & Van Oost, K. (2020). Vis-NIR spectroscopic assessment of soil aggregate stability and aggregate size distribution in the Belgian Loam Belt. *Geoderma*, 357(May 2019), 113958.
<https://doi.org/10.1016/j.geoderma.2019.113958>
- Singh, R. (2002). C. V. Raman and the Discovery of the Raman Effect. *Physics in Perspective*, 4(4), 399–420. <https://doi.org/10.1007/s000160200002>
- Sleep, B., Mason, S., Janik, L., & Mosley, L. (2022). Application of visible near-infrared absorbance spectroscopy for the determination of Soil pH and liming requirements for broad-acre agriculture. *Precision Agriculture*, 23(1), 194–218.
<https://doi.org/10.1007/s11119-021-09834-7>
- Song, X., Gao, Y., Liu, Z., Zhang, M., & Wan, Y. (2019). Development of a predictive tool for rapid assessment of soil total nitrogen in wheat-corn double cropping system with hyperspectral data. *Environmental Pollutants and Bioavailability*, 5940.
<https://doi.org/10.1080/26395940.2019.1679041>
- 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>
- Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE*, 8(6).
<https://doi.org/10.1371/journal.pone.0066409>
- Stevens, A., & Ramirez-Lopez, L. (2013). An Introduction to the prospectr package. In *2013 R Package* (0.1; Vol. 136, Issue 12, pp. 1628-a-1629).
<https://doi.org/10.1176/ajp.136.12.1628-a>
- Sun, J., Shi, S., Yang, J., Gong, W., Qiu, F., Wang, L., Du, L., & Chen, B. (2019). Wavelength selection of the multispectral lidar system for estimating leaf chlorophyll and water contents through the PROSPECT model. *Agricultural and Forest Meteorology*, 266–267(November 2018), 43–52.
<https://doi.org/10.1016/j.agrformet.2018.11.035>
- Sutton, R. S., & Barto, A. G. (2019). Reinforcement Learning, An introduction. In MIT (Ed.), 余东华 张鑫宇 孙婷 (2nd ed.). Westchester Publishing Services.
- Thenkabail, P. S., Lyon, J. G., & Huete, A. (2018). HYPERSPECTRAL REMOTE SENSING OF VEGETATION. In P. S. Thenkabail, G. J. Lyon, & A. Huete (Eds.), *Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation* (2nd ed.). CRC Press. <https://doi.org/10.1201/9781315164151>
- Thurnhofer-hemsi, K., López-rubio, E., & Molina-cabello, M. A. (2020). RADIAL BASIS FUNCTION KERNEL OPTIMIZATION FOR SUPPORT VECTOR MACHINE CLASSIFIERS. *IEEE Transactions on Neural Networks and Learning Systems*.
<https://doi.org/https://doi.org/10.48550/arXiv.2007.08233>
- Tümsavaş, Z. (2017). Application of visible and near infrared reflectance spectroscopy to predict total nitrogen in soil. *Environmental Biology*, 38(September), 409–417.
[https://doi.org/10.22438/jeb/38/5\(SI\)/GM-047](https://doi.org/10.22438/jeb/38/5(SI)/GM-047)

- Tziolas, N., Tsakiridis, N., Ogen, Y., Kalopesa, E., Ben-Dor, E., Theocharis, J., & Zalidis, G. (2020). An integrated methodology using open soil spectral libraries and Earth Observation data for soil organic carbon estimations in support of soil-related SDGs. *Remote Sensing of Environment*, 244(January), 111793. <https://doi.org/10.1016/j.rse.2020.111793>
- Uhlenbeck, G. E., & Goudsmit, S. (1976). Spinning Electrons and the Structure of Spectra. *Nature*, 2(4), 391. [https://doi.org/10.1016/0304-3762\(76\)90072-9](https://doi.org/10.1016/0304-3762(76)90072-9)
- United States Department of Agriculture. (1987). USDA Textural Soil Classification. In *Soil Mechanics Level I Module 3 - USDA Textural Soil Classification* (pp. 1–53).
- Van Reeuwijk, L. P. (2002). *Procedures For Soil Analysis* (L. P. Van Reeuwijk (ed.); 6th ed.). International Soil Reference and Information Centre.
- Vanderlinde, J., & Neuenschwander, D. E. (1994). Classical Electromagnetic Theory. In *American Journal of Physics* (Vol. 62, Issue 7). <https://doi.org/10.1119/1.17492>
- Vapnik, V. (1998). *of Function Estimation*. 55–85. https://doi.org/10.1007/978-1-4615-5703-6_5
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5), 988–999. <https://doi.org/10.1109/72.788640>
- Vestergaard, R.-J., Adamchuk, V., & Biswas, A. (2021). Evaluation of Optimized Preprocessing and Modeling Algorithms for Prediction of Soil Properties Using VIS-NIR Spectroscopy. *Sensors (Basel, Switzerland)*. <https://doi.org/doi.org/10.3390/s21206745>
- Vestergaard, R., Vasava, H., Aspinall, D., Chen, S., Gillespie, A., Adamchuk, V., & Biswas, A. (2021). *Evaluation of Optimized Preprocessing and Modeling*. 3.
- Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., ... Ji, W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155, 198–230. <https://doi.org/10.1016/j.earscirev.2016.01.012>
- Viscarra Rossel, R. A., Rizzo, R., Demattê, J. A. M., & Behrens, T. (2010). Spatial Modeling of a Soil Fertility Index using Visible-Near-Infrared Spectra and Terrain Attributes. *Soil Science Society of America Journal*, 74(4), 1293–1300. <https://doi.org/10.2136/sssaj2009.0130>
- 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(1), 88–96. <https://doi.org/10.1016/j.geoderma.2014.01.013>
- Wambugu, N., Chen, Y., Xiao, Z., Tan, K., Wei, M., Liu, X., & Li, J. (2021). Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks : A review. *International Journal of Applied Earth Observations and Geoinformation*, 105, 102603. <https://doi.org/10.1016/j.jag.2021.102603>
- Wan, M., Qu, M., Hu, W., Li, W., Zhang, C., Cheng, H., & Huang, B. (2019). Estimation of soil pH using PXRF spectrometry and Vis-NIR spectroscopy for rapid environmental risk assessment of soil heavy metals. *Process Safety and Environmental Protection*, 132, 73–81. <https://doi.org/10.1016/j.psep.2019.09.025>
- Wang, C., Liu, B., Liu, L., Zhu, Y., Hou, J., Liu, P., & Li, X. (2021). A review of deep learning used in the hyperspectral image analysis for agriculture. In *Artificial Intelligence Review* (Vol. 54, Issue 7). Springer Netherlands. <https://doi.org/10.1007/s10462-021-10018-y>
- Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root

- Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, 324(1). <https://doi.org/10.1088/1757-899X/324/1/012049>
- Warren, G. P., & Whitehead, D. C. (1988). Available soil nitrogen in relation to fractions of soil nitrogen and other soil properties. *Plant and Soil*, 112(2), 155–165. <https://doi.org/10.1007/BF02139991>
- Wartini, N., Minasny, B., Montazerolghaem, M., & Padarian, J. (2019). *Convolutional Neural Network for simultaneous prediction of several soil properties using visible / near-infrared , mid- infrared , and their combined spectra* [University of Sydney]. <https://www.sciencedirect.com/science/article/pii/S0016706119300588>
- Wehrle, R., Welp, G., & Pätzold, S. (2021). Total and hot-water extractable organic carbon and nitrogen in organic soil amendments: Their prediction using portable mid-infrared spectroscopy with support vector machines. *Agronomy*, 11(4). <https://doi.org/10.3390/agronomy11040659>
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236(December 2018), 111402. <https://doi.org/10.1016/j.rse.2019.111402>
- Wenjun, J., Zhou, S., Jingyi, H., & Shuo, L. (2014). In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy. *PLoS ONE*, 9(8). <https://doi.org/10.1371/journal.pone.0105708>
- Wilson, M. J. (2019). The importance of parent material in soil classification: A review in a historical context. *Catena*, 182(May), 104131. <https://doi.org/10.1016/j.catena.2019.104131>
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37–52. https://doi.org/10.1007/978-0-387-87811-9_2
- Wu, M., Wang, S., Pan, S., Terentis, A. C., & Strasswimmer, J. (2021). Deep learning data augmentation for Raman spectroscopy cancer tissue classification. *Scientific Reports*, 1–13. <https://doi.org/10.1038/s41598-021-02687-0>
- Xiao, S., & He, Y. (2019). Application of near-infrared spectroscopy and multiple spectral algorithms to explore the effect of soil particle sizes on soil nitrogen detection. *Molecules*, 24(13). <https://doi.org/10.3390/molecules24132486>
- Xu, G., Li, Z., & Li, P. (2013). Fractal features of soil particle-size distribution and total soil nitrogen distribution in a typical watershed in the source area of the middle Dan River, China. *Catena*, 101, 17–23. <https://doi.org/10.1016/j.catena.2012.09.013>
- Xu, S., Zhao, Y., Wang, M., & Shi, X. (2018). Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. *Geoderma*, 310(July 2017), 29–43. <https://doi.org/10.1016/j.geoderma.2017.09.013>
- Yan, X. T., Donaldson, K. M., Davidson, C. M., & Gao, Y. (2018). Effects of sample pretreatment and particle size on the determination of nitrogen in soil by portable LIBS and potential use on robotic-borne remote Martian and agricultural soil analysis systems. *RSC Advances*, 36886–36894. <https://doi.org/10.1039/C8RA07065B>
- Yang, H., Kuang, B., & Mouazen, A. M. (2012). Quantitative analysis of soil nitrogen and carbon at a farm scale using visible and near infrared spectroscopy coupled with wavelength reduction. *European Journal of Soil Science*, 63(3), 410–420. <https://doi.org/10.1111/j.1365-2389.2012.01443.x>
- Yang, Haiqing. (2011). Spectroscopic calibration for soil N and C measurement at a farm scale. *Procedia Environmental Sciences*, 10(PART A), 672–677.

- <https://doi.org/10.1016/j.proenv.2011.09.108>
- 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(January), 114616. <https://doi.org/10.1016/j.geoderma.2020.114616>
- Yang, M., Mouazen, A., Zhao, X., & Guo, X. (2020). Assessment of a soil fertility index using visible and near-infrared spectroscopy in the rice paddy region of southern China. *European Journal of Soil Science*, 71(4), 615–626. <https://doi.org/10.1111/ejss.12907>
- Yang, M., Xu, D., Chen, S., Li, H., & Shi, Z. (2019). Evaluation of machine learning approaches to predict soil organic matter and pH using vis-NIR spectra. *Sensors (Switzerland)*, 19(2). <https://doi.org/10.3390/s19020263>
- Yang, X. S. (2009). Firefly algorithms for multimodal optimization. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5792 LNCS, 169–178. https://doi.org/10.1007/978-3-642-04944-6_14
- Yao, Xiangqian, Yang, W., Li, M., Zhou, P., Chen, Y., Hao, Z., & Liu, Z. (2018). Prediction of Total Nitrogen in Soil Based on Random Frog Leaping Wavelet Neural Network. *IFAC-PapersOnLine*, 51(17), 660–665. <https://doi.org/10.1016/j.ifacol.2018.08.121>
- Yao, Xiong, Yu, K., Deng, Y., Liu, J., & Lai, Z. (2019). Spatial variability of soil organic carbon and total nitrogen in the hilly red soil region of Southern China. *Springer*.
- Young, A. T. (1981). Rayleigh scattering. *Applied Optics*, 20(4), 533–535.
- Yuan, J., Wang, X., Yan, C., Chen, S., Wang, S., Zhang, J., Xu, Z., Ju, X., Ding, N., Dong, Y., & Zhang, W. (2020). Wavelength selection for estimating soil organic matter contents through the radiative transfer model. *IEEE Access*, 8, 176286–176293. <https://doi.org/10.1109/ACCESS.2020.3026813>
- Yubing Wang, Huang, H., & Chen, X. (2021). Predicting Organic Matter Content, Total Nitrogen and pH Value of Lime Concretion Black Soil Based on Visible and Near Infrared Spectroscopy. *Eurasian Soil Science*, 54(11), 1681–1688. <https://doi.org/10.1134/S1064229321110144>
- Yun, Y. H., Li, H. D., Deng, B. C., & Cao, D. S. (2019). An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, 113, 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>
- Yun, Y. H., Wang, W. T., Tan, M. L., Liang, Y. Z., Li, H. D., Cao, D. S., Lu, H. M., & Xu, Q. S. (2014). A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Analytica Chimica Acta*, 807, 36–43. <https://doi.org/10.1016/j.aca.2013.11.032>
- Yun, Y., Li, H., Deng, B., & Cao, D. (2019). *Trends in Analytical Chemistry An overview of variable selection methods in multivariate analysis of near-infrared spectra*. 113, 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>
- Zhang, Jian, Gao, B., Chai, H., Ma, Z., & Yang, G. (2016). Identification of DNA-binding proteins using multi-features fusion and binary firefly optimization algorithm. *BMC Bioinformatics*, 17(1), 1–12. <https://doi.org/10.1186/s12859-016-1201-8>
- Zhang, Jifeng, Wang, Z., Fan, B., Hou, Y., Dou, Y., Ren, Z., & Chen, X. (2020). *Investigating the Proper Application Rate of Nitrogen under Mulched Drip Irrigation to Improve the Yield and Quality of Tomato in Saline Soil*.
- Zhang, Y., Li, M. Z., Zheng, L. H., Zhao, Y., & Pei, X. (2016). Soil nitrogen content forecasting based on real-time NIR spectroscopy. *Computers and Electronics in Agriculture*, 124, 29–36. <https://doi.org/10.1016/j.compag.2016.03.016>
- Zhang, Y., Li, M., Zheng, L., Qin, Q., & Suk, W. (2019). Spectral features extraction for

- estimation of soil total nitrogen content based on modified ant colony optimization algorithm. *Geoderma*, 333(June 2018), 23–34.
<https://doi.org/10.1016/j.geoderma.2018.07.004>
- Zhou, P., Yang, W., Li, M., Yao, X., & Liu, Z. (2018). Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Tot. *IFAC-PapersOnLine*, 51(17), 51–56. <https://doi.org/10.1016/j.ifacol.2018.08.071>
- Zhou, P., Zhang, Y., Yang, W., Li, M., Liu, Z., & Liu, X. (2019). Development and performance test of an in-situ soil total nitrogen-soil moisture detector based on near-infrared spectroscopy. *Computers and Electronics in Agriculture*, 160(January), 51–58. <https://doi.org/10.1016/j.compag.2019.03.016>

CAPITULO 3. EVALUACIÓN DE TRES MÉTODOS DE SELECCIÓN DE VARIABLES PARA PREDICCIÓN DE NITRÓGENO TOTAL Y PH DEL SUELO USANDO DATOS ESPECTRALES VIS-NIR

RESUMEN

La espectroscopia es una técnica quimiométrica utilizada para identificar de forma cualitativa, cuantitativa y no destructiva, los elementos individuales en compuestos como el suelo. En este estudio, tres algoritmos de selección de variables espectroscópicas para modelos de predicción de Nitrógeno total y pH de suelo fueron evaluados. Se exploró el desempeño del algoritmo luciérnaga binario (BFFA) y del algoritmo Genético modificado (GA) en análisis de suelos, se evaluó además el algoritmo de muestreo competitivo adaptativo reponderado (CARS). Se utilizaron datos espectrales de suelo Vis-NIR de una librería espectral de suelos de gran escala. Los espectros fueron clasificados según su clase textural, obteniendo 10 tipos. El filtro derivativo de Savitzky-Golay de segundo orden fue aplicado en los espectros como pretratamiento. Los métodos se acoplaron con PLS para calibrar los modelos. Los valores de error con validación cruzada RMSEcv, coeficiente de correlación de predicción R^2_p y la Relación de Desviación de Predicción RPD indicaron mejores desempeños de los modelos PLS con GA y BFFA en la mayoría de los subconjuntos, tanto para TN (RMSEcv= 0.73 - 2, R^2_p = 0.73 - 0.86) como para pH (R^2_p = 0.7 - 0.76, RMSEcv = 0.52 - 0.8). Se realizó un análisis de las longitudes de onda clave para cada elemento, de acuerdo con los métodos de selección empleados.

Palabras clave: Agricultura de Precisión, Algoritmos de optimización, Espectroscopía, librería espectral, Propiedades del suelo.

3.1 INTRODUCCIÓN

Se puede lograr una producción más sostenible con Agricultura de Precisión (AP) mediante decisiones respaldadas con algoritmos de aprendizaje automático y el uso de tecnologías web (Alibabaei et al., 2022).). AP se enfoca en la óptima gestión de recursos agrícolas, donde el suelo y su composición juegan un papel fundamental. El nitrógeno (N) es esencial para el crecimiento de las plantas, absorbido en forma de nitrato (NO_3^-) o amonio (NH_4^+) para la formación de aminoácidos y proteínas (Patel et al., 2020). Además, el pH del suelo, una

característica química, afecta la disponibilidad de nutrientes y los ciclos biogeoquímicos (Gondal *et al.*, 2021).

A pesar de la importancia de las propiedades del suelo, los enfoques tradicionales de análisis resultan inviables en AP debido a los recursos, equipos y costos implicados. La Espectroscopía emerge como una prometedora técnica en AP, siendo no destructiva, rápida y precisa, y eliminando la necesidad de reactivos químicos (Viscarra *et al.*, 2016). Esta técnica evalúa la reflectancia y absorción de radiación en diversas longitudes de onda (WL) del espectro electromagnético, particularmente en la región del infrarrojo cercano (NIR: 2500-700 nm) y visible (Vis: 700-350 nm), permitiendo la identificación de elementos individuales en el suelo, pues cada especie química produce un espectro de absorción único (Isaac y Na, 2017). No obstante, los espectros Vis-NIR son frecuentemente alterados y contaminados por efectos de dispersión en las muestras, que pueden afectar directamente la habilidad de predicción de modelos obtenidos a partir de ellos (Barra *et al.*, 2021). Además, la humedad (Ji, Li, *et al.*, 2016), óxidos presentes en el suelo (Warren y Whitehead, 1988) y el tamaño de partícula (Yan *et al.*, 2018) también influyen en los análisis espectrométricos. Para superar estas barreras, es necesario incorporar técnicas numéricas y estadísticas, que permitan depurar la información compleja de los espectros y correlacionarla con las propiedades del suelo a través de modelos multivariantes (Vohland *et al.*, 2014). Está demostrado que la aplicación de pretratamientos y la optimización de datos mejoran significativamente las predicciones de las propiedades del suelo (Coutinho *et al.*, 2019; Yao *et al.*, 2018; Zhang, *et al.*, 2016; Zhang *et al.*, 2019).

Los Métodos de Selección de Variables espectrales (MSV) combinan técnicas numéricas de optimización, análisis estadístico y regresión multivariante (Yun, *et al.*, 2019), para evaluar la contribución de cada variable espectral o longitud de onda al desempeño de un modelo. Entre estos enfoques, la Regresión de Mínimos Cuadrados Parciales (PLS) es ampliamente utilizada en los MSV para suelos debido a su capacidad para predecir variables dependientes a partir de un extenso conjunto de variables independientes (Qi *et al.*, 2018). Además, en el campo de las ciencias del suelo, se destacan algoritmos como los Genéticos (GA)

y el Muestreo Reponderado Adaptativo Competitivo (CARS), que se emplean con frecuencia en MSV (Barra *et al.*, 2021; Yun *et al.*, 2019), incluyendo análisis espectral para TN y pH del suelo (Y. Wang *et al.*, 2021; Nie *et al.*, 2017). Además es importante considerar versiones modificadas de estos algoritmos como el GA propuesto por Fei *et al.*, (2009), así como los resultados prometedores que han alcanzado métodos como el Algoritmo Luciérnaga en otras ramas de la ciencia (El-Zeiny *et al.*, 2021), podrían ser adecuados para su aplicación en análisis de suelos.

Es fundamental destacar que varios prototipos desarrollados para el análisis in situ del suelo mediante espectroscopía utilizan fotodetectores infrarrojos y diodos emisores como fuentes de radiación, los cuales abarcan solo algunas longitudes de onda específicas (Isaac y Na, 2017; Masrie, *et al.*, 2018; Masrie, *et al.*, 2019; Li *et al.*, 2019; Zhou *et al.*, 2019). Por lo tanto, es esencial adquirir un conocimiento más completo sobre las longitudes de onda que caracterizan a cada elemento presente en el suelo. La evaluación y selección de MSVs apropiados representan una contribución significativa al desarrollo tecnológico en Agricultura de Precisión (AP). En este contexto, las librerías espectrales de suelo a gran escala pueden servir como referencias valiosas para establecer asociaciones estadísticas sólidas (Lanfa Liu *et al.*, 2018; Tziolas *et al.*, 2020). Estas librerías caracterizan una amplia variedad de suelos para los cuales se han estandarizado los protocolos de análisis, en contraste con los enfoques utilizados en estudios locales (Stevens *et al.*, 2013). Librerías espectrales de gran escala como “Land Use Cover Area Frame Survey” (LUCAS), “The Brazilian library” o “The Chinese Library” han sido empleadas para evaluar métodos de calibración y selección, así como para apoyar estudios locales y obtener modelos más robustos (Ji *et al.*, 2016; Lei y Sun, 2022).

Por lo tanto, el objetivo de este estudio fue implementar y evaluar el desempeño de tres MSV, que incluyen una versión de GA, una versión binaria de FFA y el método CARS, para extracción de WL características y predicción de TN y pH del suelo, a partir de datos espectroscópicos disponibles en la librería espectral de suelos a gran escala LUCAS.

3.2 MATERIALES Y MÉTODOS

3.2.1 Librería Espectral LUCAS

La librería espectral LUCAS fue iniciada en 2009 como una base de datos relacional por la Unión Europea y es considerada como la más grande a escala global en propiedades de suelo (Cameron *et al.*, 2018). La librería fue actualizada en 2018 y cuenta con 19,967 muestras espectrales con resolución de 0.5nm en las regiones Vis-NIR. Las muestras fueron colectadas en la capa superficial del suelo (0-20 cm). Los análisis de las propiedades de suelo se llevaron a cabo bajo los estándares de los métodos ISO (Van Reeuwijk, 2002). Los espectros se midieron con un analizador de contenido rápido FOSS XDS (Lanfa Liu *et al.*, 2018). Para este estudio, se obtuvieron de la librería 10 subconjuntos de muestras espectrales de suelo, correspondientes a su clase textural de acuerdo con el Departamento de Agricultura de Estados Unidos (1987) como sigue: T1 = Arcilloso(662 muestras), T2 = Franco arcilloso(1324), T3 = Arcillo limoso(734), T4 = Franco arcillo-limoso (1306), T5 = Franco arcillo-arenoso (365), T6 = Franco(2965), T7 = Franco arenoso (3420), T8 = Franco limoso (2663), T9 = Arenoso franco (1979), T10 = Arenoso (1182). El 75% de los datos se destinaron para entrenamiento y el 25% para predicción.

3.2.2 Pre-Procesamiento de los Datos Espectrales

Para resaltar las características de absorción y eliminar el ruido, se utilizó el filtro derivativo de Savitzky y Golay, (1964). Esta técnica polinómica derivativa permite preservar rasgos clave de la distribución inicial, como los máximos y mínimos relativos, así como el ancho de los picos. En el presente estudio, se aplicó una transformación de los espectros de reflectancia mediante su segunda derivada (SDR), con el fin de reducir la variabilidad en la línea de base de reflectancia y realzar las características espectrales, como se ha demostrado en investigaciones previas (Hong *et al.*, 2018). Los parámetros del filtro incluyeron un polinomio móvil de segundo orden con una ventana de 11 bandas

3.2.3 Regresión PLS con el Espectro Completo FS-PLS

Como método de referencia, se calibraron modelos PLS para relacionar Nitrógeno Total y pH del suelo a partir de datos de reflectancia SDR incluyendo todas las variables del espectro. La ecuación de regresión estándar de FS-PLS es la siguiente:

$$y = \beta_i x_i + \beta_{i+1} x_{i+1} + \dots + \beta_{2500} x_{2500} + \varepsilon, \quad i = 400:0.5:2500 \quad (3.1)$$

Donde; y = variables respuesta TN y pH, x_i = valores de reflectancia en las WL; β_i = coeficientes de regresión ponderados estimados; ε = vector error. Para determinar el número óptimo de variables latentes, se realizó una validación cruzada dejando uno fuera y evitar el sobreajuste en los modelos.

3.2.4 Métodos de selección de variables espectrales

En esta sección se indican los algoritmos utilizados en este estudio. Los parámetros utilizados en cada caso se determinaron de forma empírica, buscando minimizar el error. La combinación de los Métodos de Selección de Variables (MSV) con Regresión de Mínimos Cuadrados Parciales (PLS) se corrieron 15 veces para la calibración y predicción en cada subconjunto de datos, y se presentan los valores promedio de los resultados. Todos los análisis se llevaron a cabo utilizando el software MATLAB R2020b en un equipo con 32 GB de memoria RAM y un procesador Intel Core i7.

Algoritmo Genético. El algoritmo genético (GA) utilizado en este estudio se basa en el enfoque convencional descrito por Arcos *et al.*, (1997), pero se realizaron modificaciones en los operadores de mutación y cruzamiento. Se mantuvo constante el número de variables seleccionadas en cada generación, siguiendo la recomendación de Fei *et al.*, (2009). lo que ha demostrado mejorar la detección espectroscópica. Las modificaciones se basaron en segmentaciones dinámicas multipunto de los cromosomas, asegurando que el número de soluciones en los individuos no se alterara. En este estudio, cada individuo o cromosoma (Cr) en la población es un vector binario de dimensión n , donde n es igual al número total de variables en un espectro, es decir, $Cr \in R^n$. La codificación en cada

cromosoma para inclusión o para exclusión de la i -ésima variables espectral en la resolución del problema es $Cr(i) = 1$ y $Cr(i) = 0$ respectivamente.

Los parámetros iniciales del algoritmo son los siguientes: número de cromosomas = 50, constante de variables seleccionadas = 10, $\beta = 0.75\%$, $\alpha = 0.01\%$ y $T_{max} = 1000$.

Algoritmo Luciérnaga Binario BFFA. El Algoritmo Luciérnaga (FFA) fue desarrollado por Yang (2009), es una generalización de SPO y fue diseñado para problemas continuos. Posteriormente, surgieron las variantes binarias del FFA, conocidas como BFFA. En este estudio, se implementó el algoritmo BFFA, descrito por Zhang *et al.*, (2016). Está inspirado en el comportamiento de las luciérnagas en reproducción, utiliza la intensidad de la luminiscencia como factor intrínseco de atracción de las luciérnagas, donde esta atracción se intensifica de manera inversamente proporcional a la disminución del error. Cada luciérnaga representa una solución potencial con un subconjunto de variables seleccionadas aleatoriamente, las cuales se utilizan para calibrar un modelo PLS. La luminiscencia se evalúa a través del RMSE en toda la población, y las luciérnagas se atraen entre sí en función de su brillo relativo y distancia. De esta manera, el desplazamiento de un individuo x_i^t hacia otro con mayor luminiscencia x_j^t está dado por la ecuación 3.2:

$$X_i^{t+1} = X_i^t + \beta_0 e^{-\gamma r_{ij}^2} (x_j^t - x_i^t) + \alpha \epsilon_i^t \quad (3.2)$$

En donde, el término X_i^t guarda la posición actual del individuo i . β_0 representa la atracción entre individuos en función de su luminiscencia. La cual decrece con la distancia entre las dos luciérnagas $(x_j^t - x_i^t)$ a una tasa determinada por $e^{-\gamma r_{ij}^2}$, r es la distancia de Haming para dos vectores, γ controla la atenuación de luz y visibilidad. El tercer termino representa la aleatoriedad del movimiento de la luciérnaga, con un factor de escala α para controlar el paso del desplazamiento y una distribución aleatoria ϵ_i^t .

Si una luciérnaga es atraída por otra, sus bits cambian en función de la luciérnaga más brillante, y se permite la mutación de bits en la nueva luciérnaga. Este

proceso se repite hasta alcanzar un número predefinido de iteraciones, y al final se selecciona la luciérnaga más brillante como la solución óptima. Los parámetros iniciales fueron: máximo de iteraciones $T_{max} = 500$, número de luciérnagas = 50, $\beta_0 = 0.3$, número inicial de variables en cada luciérnaga $Nb = 25$.

Algoritmo de Muestreo Reponderado Adaptativo Competitivo CARS. Este método ha demostrado ser efectivo para la selección de variables espectroscópicas en diversos estudios relacionados con las ciencias del suelo (Yuan *et al.*, 2020; Jifeng Zhang *et al.*, 2020). El método imita el principio de "supervivencia del más apto", eliminando gradualmente las longitudes de onda invariables en el análisis. Utiliza muestreo de Montecarlo en la generación de su población, construyendo subconjuntos de variables, con los cuales se realiza una regresión PLS para obtener los coeficientes de regresión β . Después, se evalúan los pesos de las variables espectrales utilizando la magnitud de β . La selección de variables se basa en la función exponencial decreciente FED (3.3) que reduce el espacio de búsqueda en cada iteración a través de un umbral de retención de variables.

$$FED_i = \frac{p^{\frac{1}{n-1}}}{2} e^{-i \left(\frac{\ln(\frac{p}{2})}{n-1} \right)}, i = 1:M \quad (3.3)$$

Donde p = número total de variables en el espectro.

Este enfoque busca identificar de manera eficiente las longitudes de onda más relevantes en el análisis de datos espectrales. En este estudio, el número de iteraciones fue $M=50$.

3.2.5 Evaluación de los Modelos de Predicción.

La capacidad predictiva de los modelos se evaluó calculando el coeficiente de determinación de predicción (R_p^2), la raíz del error cuadrado medio RMSE (Wang y Lu, 2018) con validación cruzada dejando uno fuera (LOO) y la Desviación Predictiva Residual (RPD). El RPD se define como el cociente entre la desviación estándar de los datos de referencia y el RMSE (Bellon-Maurel *et al.*, 2010). Se

establece que valores de RPD entre 2 y 3 indican un modelo con una buena capacidad de predicción, modelos con $1.5 < \text{RPD} < 2$ son adecuados pero perfectibles, un $\text{RPD} < 1.5$ indica que el modelo tiene una pobre capacidad de predicción (D'Acqui *et al.*, 2010).

Para determinar significancia de las longitudes de onda utilizadas en los modelos, se empleó la Importancia de las Variables en la Proyección VIP (Vohland *et al.*, 2014). VIP refleja la influencia de un predictor (WL) para una variable y observada, una puntuación VIP alta (> 1) indica una variable significativa (Chong y Jun, 2005).

3.3 RESULTADOS Y DISCUSIÓN

3.3.1 TN y pH del suelo en la librería LUCAS

Se tiene en la librería un contenido medio de TN que va de 1.7a 2.8 g kg⁻¹ como se muestra en el Cuadro 3.1. Es importante destacar que se observan índices de variación que superan el 100% en suelos de textura gruesa como T9 y T10, y coeficientes de asimetría positivos con valores que van de 2 a valores mayores que 8.

Cuadro 3.1. Estadísticas de TN y pH del suelo en la librería LUCAS

	Clase									
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
TN	g/ kg									
$\bar{X}$	2.84	2.38	2.78	2.53	2.05	2.4	2.01	1.9	1.78	1.46
s_x	2.31	1.9	1.8	1.6	1.78	2.02	1.79	1.5	2.03	1.86
γ_1	3.86	2.66	2.04	2.57	2.78	6.4	6.5	5.48	7.13	8.17
CV	81.5	79.4	67.7	63	86.6	84.4	88.8	78.3	114	127
pH										
$\bar{X}$	7.034	7.24	7.15	7.0	7.11	6.46	5.98	6.57	5.74	5.31
s_x	0.99	0.99	1.02	1.05	1.01	1.12	1.19	1.12	1.06	1.04
γ_1	-0.81	-1.05	-0.90	-0.80	-0.67	-0.34	-0.04	-0.42	0.23	0.74
CV	14	13.7	14.2	15	14.2	17.2	19.9	17	18.6	19.5

$\bar{X}$ = Media aritmética, s_x = Desviación estándar, γ_1 = Coeficiente de asimetría de Fisher, CV= Coeficiente de Variación (%).

En cuanto al pH, las medias oscilan entre 5.1 y 7.2, los coeficientes de asimetría son negativos entre -0.04 y -1.05 en suelos T1-T8, y positivos en suelos de

textura más gruesa que no superan los 0.75. Asimismo, se observan coeficientes de variación por encima de 13% hasta valores cercanos al 20%.

Los indicativos de TN y pH en la librería sugieren una diversidad de suelos con variaciones en la fertilidad y la acidez/alcalinidad, pero en su mayoría dentro de rangos aceptables para la agricultura.

Tanto los valores de pH como de TN son consistentes con los reportados en otras librerías, como se evidencia en los estudios de Ji *et al.* (2016), Viscarra *et al.* (2010) y Shepherd y Walsh (2002). Sin embargo, la presencia de valores atípicos en la concentración de TN y coeficientes de asimetría altos indican que algunos suelos pueden requerir una atención especial en términos de gestión agrícola.

3.3.2 Espectros de Reflectancia

En la Figura 3.1, se muestra los valores de espectro de reflectancia promedios de cada clase textural. Se observa una disminución en la reflectancia en suelos con un contenido de limo superior al 40%, especialmente en las clases T8, T3 y T4. Esta relación entre la textura del suelo y la reflectancia tiene implicaciones significativas para la detección y caracterización de diferentes tipos de suelos en aplicaciones de teledetección y agricultura de precisión (Shi *et al.*, 2020).

Los suelos con una textura Franco-limosa mostraron consistentemente la reflectancia media más baja en todo el espectro NIR. Además, se identificaron picos de reflectancia notables alrededor de 1400 y 2200 nm, los cuales han sido previamente asociados con la estimación de TN y pH en suelos (Isaac y Na, 2017; MacAbiog *et al.*, 2020; Yao Zhang *et al.*, 2016). Asimismo, alrededor de 1900 nm se observa otro pico de reflectancia que también puede proporcionar información relevante para la estimación precisa del pH, según estudios anteriores (Mammadov *et al.*, 2022; Wan *et al.*, 2019).

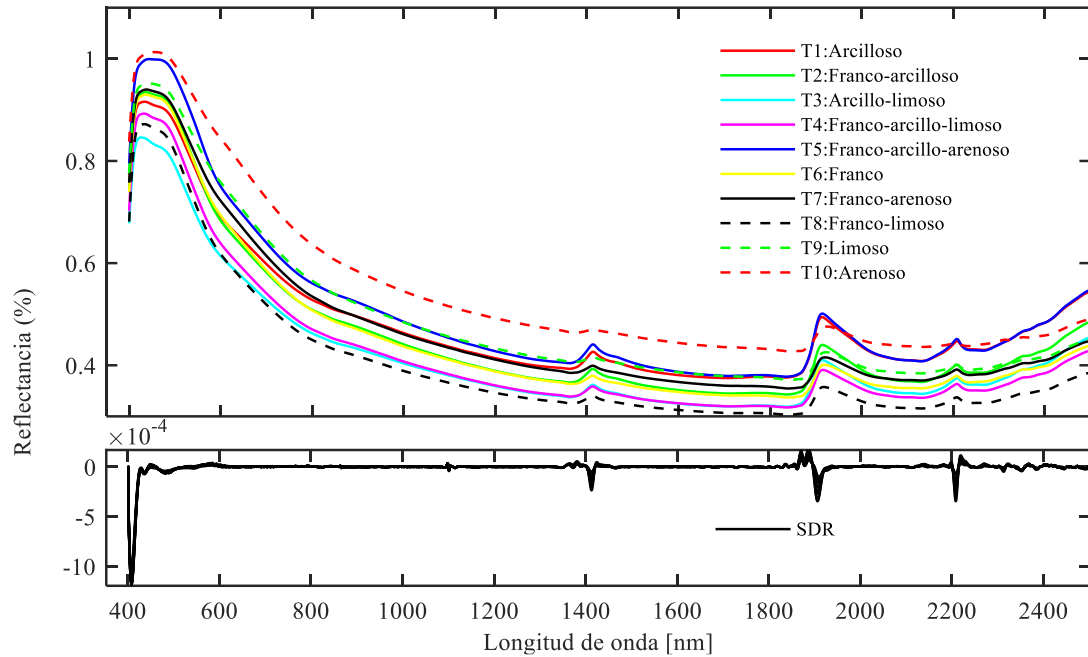


Figura 3.1. Espectros de origen promedio de cada clase textural y su segunda derivada SDR.

3.3.3 Desempeño de los métodos GA-PLS, BFFA-PLS, CARS-PLS y FS-PLS

En el Cuadro 3.2 se presentan los parámetros de predicción R_p^2 , el $RMSE_{cv}$ y RPD de los modelos PLS con el 100 % de las variables incluidas (espectro completo FS-PLS). Se observó que los modelos tuvieron un mejor desempeño en la predicción de los niveles de TN en suelos con texturas de carácter medio o franco. Sin embargo, vale la pena destacar que, aunque se lograron los mejores ajustes en suelos T2, T6 y T8, en ningún caso se alcanzó un valor de RPD igual o superior a 2, lo que indica que la capacidad predictiva para el TN podría beneficiarse de una optimización adicional.

Por otro lado, en cuanto al pH, los índices de ajuste R_p^2 presentaron valores muy similares en todos los tipos de suelos, con la mayor diferencia entre los suelos T3 y T4. Lo que sugieren que la predicción del pH se mantuvo relativamente consistente en todos los suelos analizados, independientemente de su textura. Se observan errores menores que para TN, lo cual se puede relacionar a la variación elevada en los datos de TN (Cuadro 3.1).

Cuadro 3.2. Parámetros de predicción de TN y pH del suelo con modelos de espectro completo FS-PLS en cada clase de suelo.

Clase	TN			pH		
	RMSEcv	R ²	RPD	RMSEcv	R ²	RPD
T1	1.16	0.69	1.63	0.53	0.73	1.84
T2	0.96	0.79	1.98	0.63	0.73	1.58
T3	1.46	0.81	1.21	0.90	0.78	1.24
T4	1.41	0.85	1.18	0.79	0.65	1.42
T5	0.76	0.83	1.79	0.56	0.74	1.76
T6	0.90	0.83	1.81	0.64	0.73	1.45
T7	1.98	0.75	1.25	0.78	0.70	1.30
T8	0.94	0.76	1.88	0.71	0.73	1.32
T9	0.94	0.71	1.18	0.76	0.72	1.27
T10	1.71	0.70	1.13	0.86	0.75	0.94

RMSEcv=Error cuadrado medio de validación cruzada (LOO)

En el Cuadro 3.3 se muestran los parámetros predicción de los modelos PLS con cada MSV para cada clase. La eliminación de WL redundantes o no relacionadas mejoraron enormemente la parsimonia de los modelos respecto a FS-PLS. En la mayoría de los suelos, los modelos de predicción con un número reducido de variables espectrales tuvieron un menor RMSEcv en comparación con los modelos FS-PLS, lo que resultó en una mejora en la capacidad predictiva RPD. Reafirmado que el rendimiento de los modelos PLS se puede mejorar mediante la selección de WL(H. Li *et al.*, 2022; Yang *et al.*, 2020). El porcentaje de información retenida por los métodos respecto al espectro completo fue el siguiente: GA utilizó sólo el 0.23% del espectro, pues en cada modelo se incluyeron únicamente 10 variables espectrales, tanto para TN como para pH; BFFA utilizó en promedio más información para TN que para pH, con el 1.25% y 0.98% respectivamente; CARS utilizó un 1% del espectro para TN y fue el método que más información utilizó para pH empleando más del 4.4 % en promedio. Estos porcentajes sugieren que al menos un 90% de la información de los espectros de reflectancia estudiados podría ser redundante en la predicción de TN y pH en suelos.

Particularmente para TN, los métodos GA-PLS y BFFA-PLS (Figuras 3.2a y 3.2b) se destacaron como los mejores en suelos T1, T3, T4 y T8, mientras que en T2 y T6, BFFA-PLS mostró un buen desempeño. Los mejores resultados con CARS-

PLS se obtuvieron en suelos T6 (Figura 3.2c), aunque no superaron los parámetros promedio de GA y BFFA.

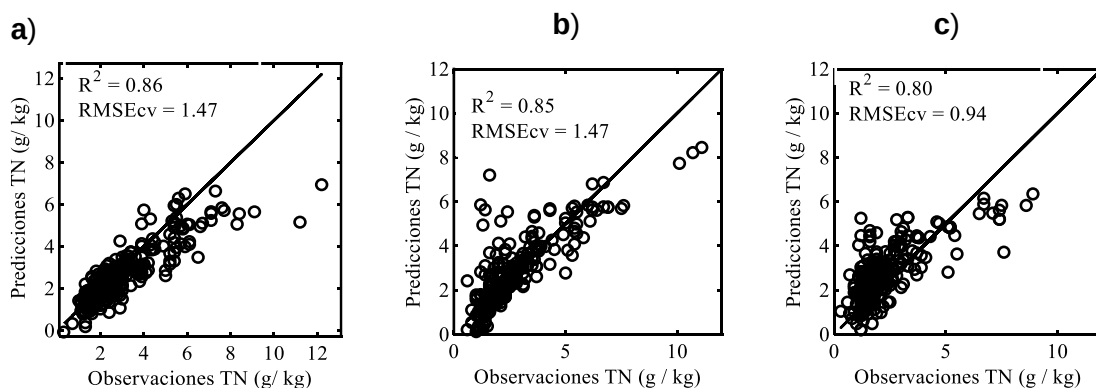
Cuadro 3.3. Parámetros de predicción de TN y pH con modelos GS-PLS, BFFA-PLS y CARS-PLS en cada tipo de suelo.

Clas e	MSV	TN				pH			
		RMSE t	RMSE p	R ²	RP D	RMSE t	RMSE p	R ²	RP D
T1	GA	1.12	1.07	0.78	1.81	0.49	0.52	0.74	1.90
	BFFA	1.01	1.10	0.76	1.92	0.59	0.61	0.66	1.64
	CARS	1.78	1.44	0.57	1.34	0.79	0.64	0.50	1.54
T2	GA	0.83	0.89	0.80	2.12	0.42	0.72	0.77	1.39
	BFFA	0.82	0.97	0.82	1.95	0.45	0.63	0.75	1.57
	CARS	1.06	1.13	0.70	1.68	0.71	1.02	0.38	0.97
T3	GA	0.81	0.92	0.82	1.93	0.47	1.04	0.77	1.08
	BFFA	0.84	1.54	0.79	1.15	0.46	0.61	0.77	1.84
	CARS	1.10	1.55	0.68	1.15	0.75	0.79	0.50	1.43
T4	GA	0.54	1.47	0.86	1.14	0.44	0.80	0.71	1.41
	BFFA	0.58	1.41	0.84	1.18	0.71	0.85	0.69	1.33
	CARS	0.85	1.43	0.66	1.17	0.69	1.07	0.31	1.05
T5	GA	0.76	0.83	0.83	1.65	0.49	0.56	0.76	1.76
	BFFA	0.86	0.79	0.79	1.72	0.66	0.61	0.71	1.62
	CARS	1.50	1.04	0.56	1.35	0.96	0.53	0.58	1.85
T6	GA	0.63	0.89	0.81	1.83	0.53	0.56	0.76	1.59
	BFFA	0.67	0.84	0.78	1.94	0.55	0.58	0.75	1.71
	CARS	0.80	0.99	0.73	1.65	0.89	0.75	0.66	1.26
T7	GA	0.60	2.02	0.79	1.22	0.61	0.79	0.71	1.28

	BFFA	0.67	2.00	0.7 3	1.24	0.64	0.76	0.6 8	1.33
	CAR S	0.76	2.11	0.6 8	1.19	0.94	0.86	0.6 4	1.26
	GA	0.53	1.09	0.8 3	1.62	0.50	0.66	0.7 4	11.5
T8	BFFA	0.65	0.97	0.7 4	1.82	0.50	0.79	0.7 3	1.56
	CAR S	0.84	1.15	0.6 1	1.57	0.81	0.80	0.6 5	1.26
	GA	1.23	1.12	0.8 3	1.62	0.63	0.78	0.7 4	11.5
T9	BFFA	1.38	0.97	0.7 4	1.82	0.59	0.61	0.7 3	1.56
	CAR S	1,83	1.13	0.6 1	1.57	0.95	0.92	0.6 5	1.26
	GA	1.20	1.46	0.7 5	1.33	0.62	0.79	0.7 1	1.03
T10	BFFA	1.43	1.73	0.7 2	1.12	0.62	0.87	0.7 2	0.93
	CAR S	1,78	1.55	0.4 8	1.26	0.93	0.85	0.4 0	0.94

RMSEt y RMSEp: Errores cuadrados medios de entrenamiento y predicción respectivamente con validación cruzada (LOO).

Con respecto a otros estudios con la misma librería (Mendes *et al.*, 2021), en dichos suelos se mejoró la precisión de predicción de los modelos PLS con GA y BFFA para TN. Sin embargo, los resultados con los tres MSV en T4, T7, T9 y T10 fueron poco consistentes, con índices de rendimiento (RPD) por debajo de 1.5 (Figura 3.3a), lo que sugiere deficiencias en la predicción en suelos franco-arcillosos, franco-limosos, areno francos y arenosos, debido a la presencia de altos porcentajes de arena, que dificultan la detección de TN, de acuerdo con estudios previos como el de Xiao y He (2019).



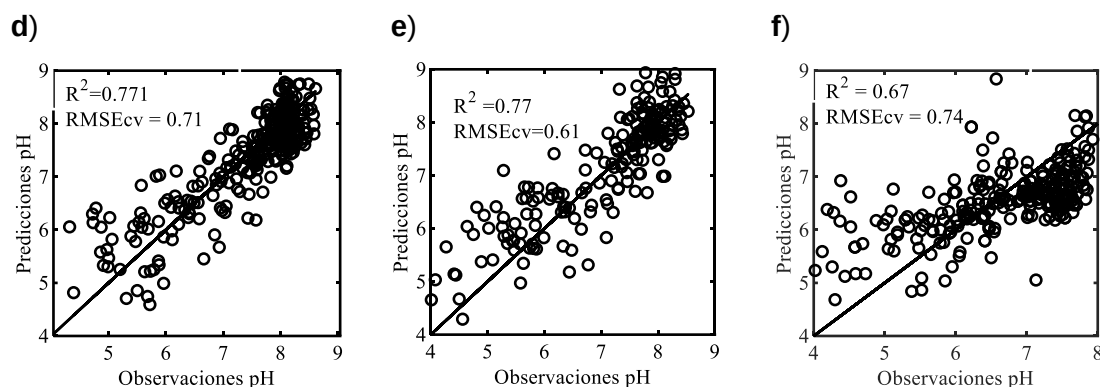


Figura 3.2. Observaciones y predicciones de los mejores modelos PLS con cada MSV. Secciones a, b y c; TN en T4, T2 y T6 con GA, BFFA y CARS respectivamente. Secciones d, e y f; pH en T5, T3 y T6 con GA, BFFA y CARS respectivamente. Se indica el coeficiente R^2 de predicción y el RMSEcv (LOO).

Para pH se obtuvieron predicciones más precisas con GA-PLS y BFFA-PLS (Figuras 3.2d y 3.2e). En T1 y T5, GA-PLS alcanzó valores de R^2_p superiores y RMSEcv mínimos, con RPD de 1.9 y 1.76 respectivamente (Figura 3.3b). En T3 y T6, BFFA-PLS se desempeñó mejor con RPD de 1.84 y 1.71. Los mejores rendimientos de CARS se observaron en suelos T5 y T6 (Figura 3.2f).

Estas diferencias pueden atribuirse a las diferencias que existen en los principios fundamentales de cada método (Yuan *et al.*, 2020). Es decir, la selección de variables en CARS depende de los parámetros de regresión de PLS, mientras que GA y BFFA pertenecen a los algoritmos evolutivos (Eiben, 2015). Esto implica el uso de estrategias de optimización inteligente basados en conceptos naturales de adaptación, competencia y reproducción, como los operadores de mutación y cruzamiento en GA y el desplazamiento en BFFA. Lo que implica una retroalimentación durante el proceso de optimización. En cada iteración, GA y BFFA generan nuevos individuos relacionados con soluciones potencialmente exitosas (predecesores), CARS al ser un método de naturaleza aleatoria constante (Vohland *et al.*, 2014) carece de este tipo de retroalimentación. La diferencia en el número de variables seleccionadas también se relaciona con estas distinciones.

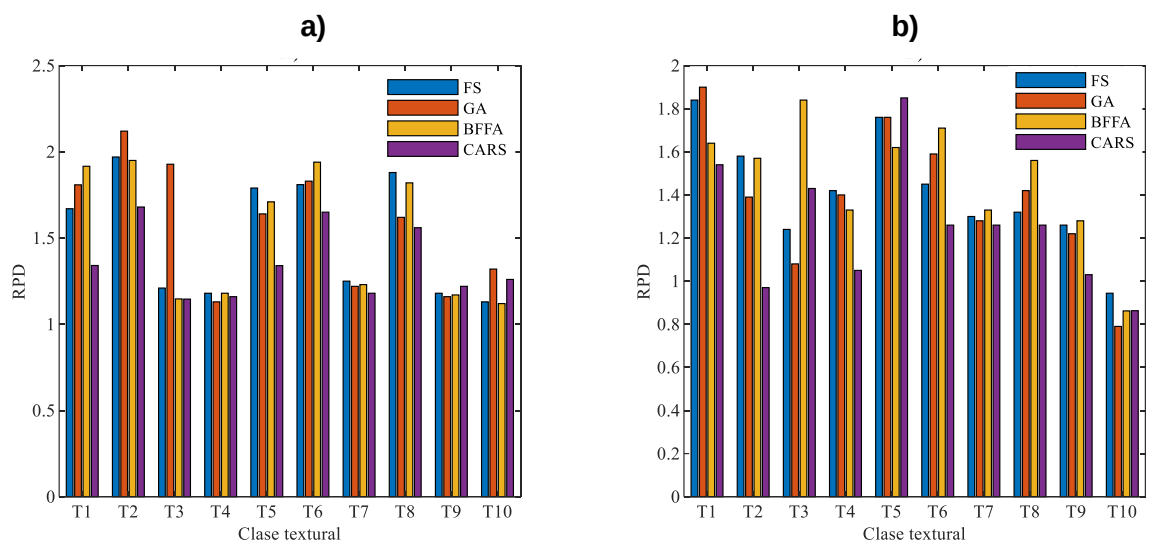


Figura 3.3. Desempeño de los modelos en función del RPD para la predicción de TN (sección a) y pH (sección b) para cada subconjunto de datos.

Es importante destacar que mayoría de modelos no supero un RPD mayor a 2, lo que está asociado según Stevens *et al.* (2013) a que los suelos son materiales complejos con una fuerte estructura espacial y la relación que vincula las propiedades del suelo con los espectros no es estacionaria ni exclusivamente lineal. Además, el desempeño de estos métodos está directamente relacionado con la calidad y representatividad de los datos de entrenamiento (Sun *et al.*, 2019). La asimetría positiva en TN y negativa en pH (Cuadro 3.1) induce un gran sesgo en los modelos cuando se presentan contenidos altos de TN y valores de pH bajos, lo cual se puede observar en los ajustes de predicción (Figura 3.2).

3.3.4 Longitudes de onda seleccionadas por los MVS

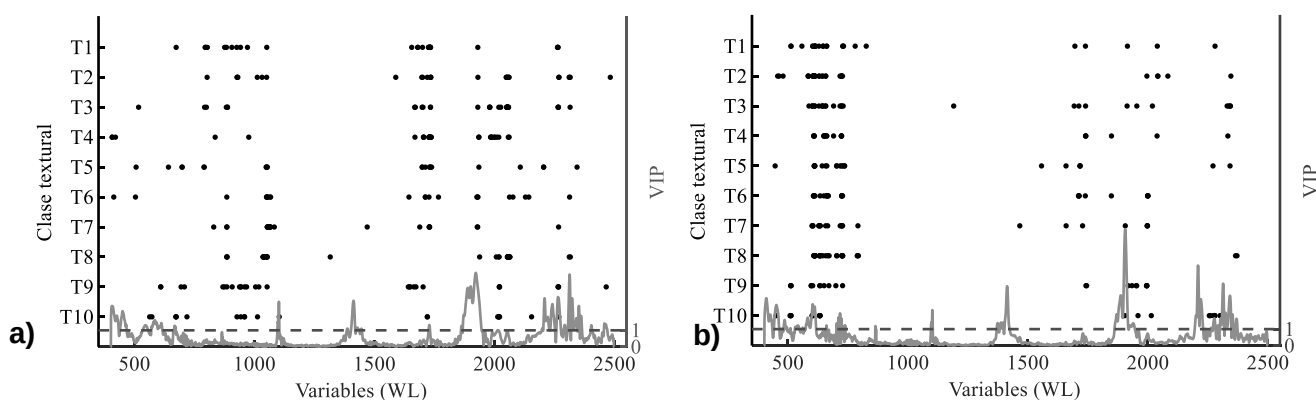
Las WL seleccionadas por los MVS para estimaciones de TN y pH en cada tipo de suelo se muestran en la Figura 3.4, incluyendo valores VIP de cada variable espectral. Para TN, los patrones de selección de GA (Figura 4a), se centraron en las variables alrededor de 1100, 1727, 2260-2313 nm (VIP > 1) anteriormente identificadas (Li *et al.*, 2019; Yao *et al.*, 2018). Se observaron también frecuencias de selección en zonas cercanas a 870-1055 nm para la mayoría de los suelos, estas WL también han sido descritas como indicadores de TN (Mukherjee y Laskar, 2019; Yao Zhang *et al.*, 2019). Así mismo, las WL cercanas a 1315 y 1469 nm reportadas por Zhou *et al.*, (2018), fueron importantes para

las clases T8 y T7. Así mismo, WL menos exploradas como la región de 2465-2482 nm fueron incluidas para clases T9 y T2.

La selección hecha por CARS-PLS para TN (Figura 3.4c) coincide con una puntuación VIP >1 en la mayoría de las regiones. Las WL próximas a 404-466, 1885-1940 y 2305-2364 nm fueron importantes para todas las clases, ya que son regiones características para Nitrógeno en suelos (Masrie *et al.*, 2018, 2019; Zhang *et al.*, 2016). Al rededor de 1415 nm es una zona de absorción conocida en estimaciones de TN (P. C. Nie *et al.*, 2017). Para la mayoría de los suelos esta región fue informativa, a excepción de T9 y T10.

Al igual que CARS en 1850 -1927nm y a GA en 2266-2391 nm, BFFA incluyo estas WL en la mayoría de las clases texturales para TN (Figura 3.4e). Las WL de 1382, 1428, 1547 y 1668 nm, fueron particularmente importantes en modelos para clases T3, T4, T7 y T10 respectivamente.

Tanto BFFA como CARS presentaron mayor dificultad para identificar WL clave para TN en suelos con tamaño de partícula grande, principalmente en clases T9 y T10. De acuerdo con Yan *et al.*, (2018) esto sería causado por una pobre absorción en suelos con partículas de mayor tamaño. Además, BFFA eliminó la zona espectral alrededor de 1000 nm de la mayoría de los subconjuntos. Lo anterior contribuye a mejores relaciones, pues esta zona se caracteriza por la presencia de óxidos de hierro ferroso y férrico que son atenuantes de reflectividad. Cabe mencionar que las WL de 2100 a 2500 nm se seleccionaron para TN en todos los tipos de suelos con cada MSV, dicha zona tiene fuertes relaciones con el TN en suelos según Kawamura *et al.*, (2021) y H. Yang, (2011).



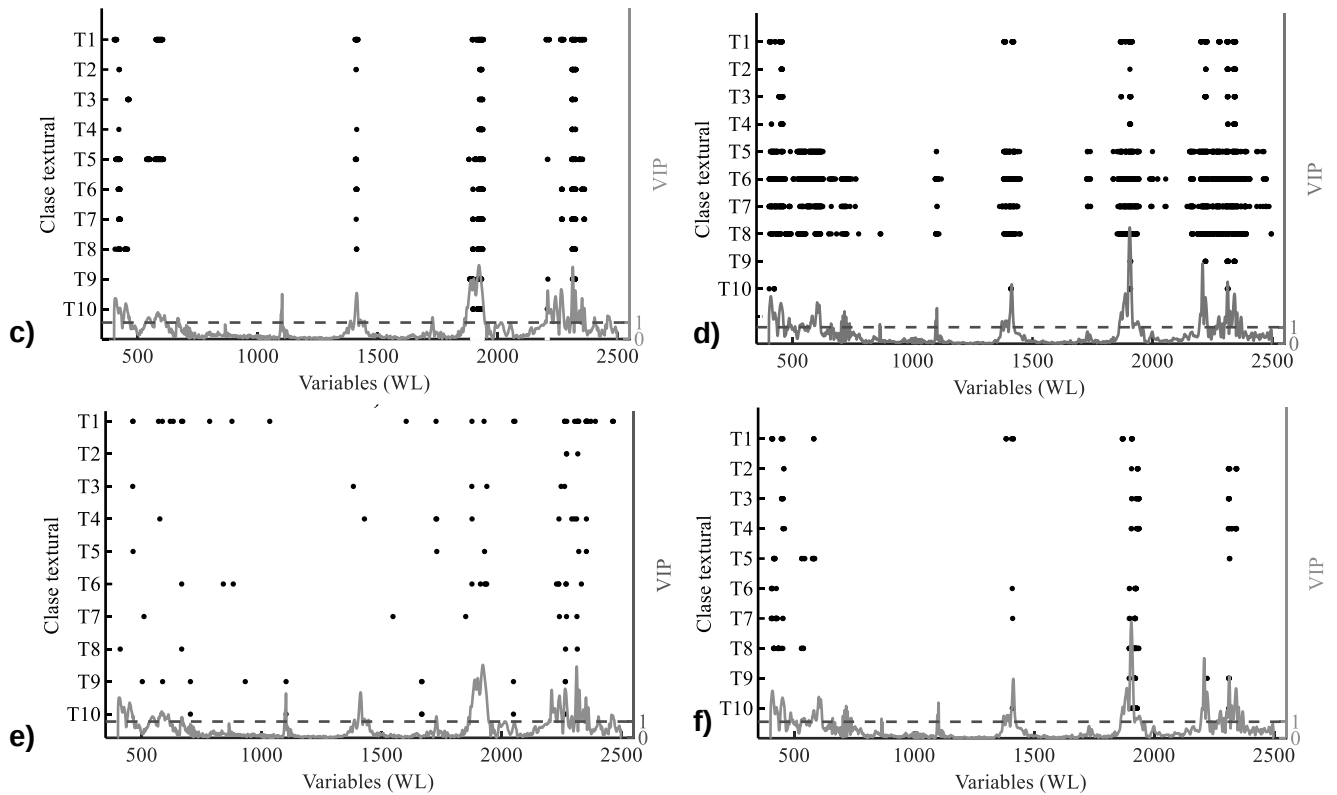


Figura 3.4. Regiones espectrales seleccionadas (ennegrecidas) por GA, CARS y BFFA para TN (secciones a, c y e respectivamente) y para pH del suelo (secciones b, d y f).

Para pH, las regiones cercanas a 598 nm, 661-713 nm y 1744 nm fueron consideradas para la mayoría de las clases texturales por GA (Figura 4 sección b). Así mismo, las WL en 2260-2371nm fueron variables particularmente importantes para los modelos en T1, T2, T3, T4, T5, T6, T8 y T10. Adicionalmente, se identificaron zonas informativas frecuentes para pH en 1894-1958 nm y 2260-2371nm, según Mammadov *et al.*,(2022).

Por otro lado, CARS determinó importantes a las regiones 1862-1944 y 2196-2372 nm en todas las clases (Figura 4 sección d). Estas regiones coinciden con las reportadas por Mammadov *et al.*, (2022) y Wan *et al.*, (2019). Además, fueron informativas las WL de 400-724 nm para el pH en suelos de clase T5, T6, T7 y T8. Las WL en 404-457 nm (Sarathjith *et al.*, 2016), fueron frecuentemente incluidas en los modelos para T1, T2, T3, T4 y T10. Las WL cercanas a 1101 nm ($Vip > 1$), fueron de importancia para modelos en suelos de clase T5, T6, T7 y

T8. Similarmente, en la zona de 1411 nm, se encontraron variables importantes para 6 de las 10 clases.

Finamente, BFFA determino diferentes zonas informativas de pH (Figura 4 sección f) en la región visible, se muestra consistente con los picos VIP altos en su selección de WL para todas las clases texturales. Las regiones NIR de importancia para la mayoría de los suelos se encuentran en el rango 1895-1936 nm y 2312-2344 nm. Además, a excepción de los suelos T10 y T12, las WL en 417-539 nm se consideraron informativas para las demás clases. Por último, las WL en 1384-1413 nm destacaron para suelos T1, T7 T8 y T12, lo cual es consistentes con reportes anteriores (Mammadov *et al.*, 2022; Sarathjith *et al.*, 2016; Wan *et al.*, 2019).

3.4 CONCLUSIONES

El desempeño predictivo de los modelos PLS se optimizó con los MSV respecto a los modelos con FS. La selección de un subconjunto de WL informativas y la eliminación de aquellas no relacionadas mejoraron enormemente la parsimonia de los modelos en comparación con FS-PLS.

En general, GA-PLS fue el algoritmo de mejor desempeño para ambos elementos en la mayoría de las clases texturales. A diferencia de CARS y BFFA, su selección de WL se propaga a través de todo el espectro Vis-NIR, lo cual lo hace un algoritmo más exploratorio a pesar de emplear la menor cantidad de información del espectro. Además, fue el más adecuado para detección de TN y pH en suelos de textura media y gruesa. Así mismo, se destaca el rendimiento y estabilidad de BFFA-PLS sobre el desempeño de CARS en las diferentes clases texturales, por lo tanto, BFFA es un método aplicable para la detección de TN y pH en muestras de suelo con espectroscopia Vis-NIR.

Los mejores resultados de predicción para TN se obtuvieron con GA-PLS y BFFA-PLS, en suelos de clases T2 y T3 y clases T1, T6 y T8 respectivamente. Para pH, GA-PLS fue mejor en T1, CARS-PLS en T5 y BFFA-PLS en T2, T3, T6 y T8.

Las WL con más frecuencia de selección para TN se encontraron en los rangos espectrales NIR: 1382, 1428, 1547, 1668, 1850 -1927, 2266-2391, y para pH en la región visible: 417-539 y en el NIR en:1744, 1895-1936, 2312-2344 y 1384-1413 nm para la mayoría de las clases texturales de suelo.

Referencias del capítulo

- Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1). <https://doi.org/10.1016/j.jjime.2020.100004>
- Ahmadi, A., Emami, M., Daccache, A., & He, L. (2021). Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis. *Agronomy*, 11(3). <https://doi.org/10.3390/agronomy11030433>
- Albawi, S., & Mohammed, T. A. (2017). Understanding of a Convolutional Neural Network. *IEEE*. <https://ieeexplore.ieee.org/abstract/document/8308186>
- Algamil, Z. Y. (2019). Variable selection in count data regression model based on firefly algorithm. *Statistics, Optimization and Information Computing*, 7(2), 520–529. <https://doi.org/10.19139/soic.v7i2.566>
- Alibabaei, K., Gaspar, P. D., Lima, T. M., Campos, R. M., Girão, I., Monteiro, J., & Lopes, C. M. (2022). A Review of the Challenges of Using Deep Learning Algorithms to Support Decision-Making in Agricultural Activities. *Remote Sensing*, 14(3), 1–43. <https://doi.org/10.3390/rs14030638>
- Andrade, R., Faria, W. M., Silva, S. H. G., Chakraborty, S., Weindorf, D. C., Mesquita, L. F., Guilherme, L. R. G., & Curi, N. (2020). Prediction of soil fertility via portable X-ray fluorescence (pXRF) spectrometry and soil texture in the Brazilian Coastal Plains. *Geoderma*, 357(July 2019), 113960. <https://doi.org/10.1016/j.geoderma.2019.113960>
- Araújo, S. R., Wetterlind, J., Demattê, J. A. M., & Stenberg, B. (2014). Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. *European Journal of Soil Science*, 65(5), 718–729. <https://doi.org/10.1111/ejss.12165>
- Arcos, J. M., Cruz Ortiz, M., Villahoz, B., & Sarabia, L. A. (1997). Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: Application on copper and zinc mixture. *Talanta*, 59(2), 311–317. [https://doi.org/10.1016/S0039-9140\(02\)00505-2](https://doi.org/10.1016/S0039-9140(02)00505-2)
- Armenise, E., Redmile-Gordon, M. A., Stellacci, A. M., Ciccicarese, A., & Rubino, P. (2013). Developing a soil quality index to compare soil fitness for agricultural use under different managements in the mediterranean environment. *Soil and Tillage Research*, 130, 91–98. <https://doi.org/10.1016/j.still.2013.02.013>
- Atasoy, S., Mateus, D., Georgiou, A., Navab, N., & Yang, G. Z. (2011). Wave interference for pattern description. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6493 LNCS(PART 2), 41–54. https://doi.org/10.1007/978-3-642-19309-5_4
- Baek, J. Y., Yoo, Y. S., & Bae, S. H. (2019). Adversarial Learning With Knowledge of Image Classification for Improving GANs. *IEEE Access*, 7, 56591–56605. <https://doi.org/10.1109/ACCESS.2019.2913697>
- Barber, D. (2020). *Bayesian Reasoning and Machine Learning*.
- Barefoot, A., Murphy, J., & Aizawa, H. (2003). *Handbook of Residue Analytical Methods for Agrochemicals* (P. W. Lee (ed.); Wiley).
- 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>
- Bartholdi, E., & Ernst, R. R. (1973). Fourier spectroscopy and the causality principle. *Journal of Magnetic Resonance* (1969), 11(1), 9–19. [https://doi.org/10.1016/0022-2364\(73\)90076-0](https://doi.org/10.1016/0022-2364(73)90076-0)
- Bellon-Maurel, V., Fernandez-ahumada, E., Palagos, B., Roger, J., & Mcbratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry*, 29(9), 1073–1081.
<https://doi.org/10.1016/j.trac.2010.05.006>
- Ben-Dor, E., Chabrilat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., & Sommer, S. (2009). Using Imaging Spectroscopy to study soil properties. *Remote Sensing of Environment*, 113(SUPPL. 1), S38–S55. <https://doi.org/10.1016/j.rse.2008.09.019>
- Ben-dor, E., & Epema, G. F. (2014). *Soil Reflectance. January 1999*.
- Ben Dor, E., Ong, C., & Lau, I. C. (2015). Reflectance measurements of soils in the laboratory: Standards and protocols. *Geoderma*, 245–246, 112–124. <https://doi.org/10.1016/j.geoderma.2015.01.002>
- Brame, C. J. (2007). A guide To active learning -Vanderbilt white paper. *Vanderbilt University Center for Teaching*. <https://cft.vanderbilt.edu/active-learning/>
- Brieman, L. (1996). Bagging predictors. *Machine Learning*, 24(3), 123–140.
<https://doi.org/10.3390/risks8030083>
- Bronick, C. J., & Lal, R. (2021). Soil structure and management : a review. *Geoderma*, 124(2005), 3–22.
<https://doi.org/10.1016/j.geoderma.2004.03.005>
- Brunton, S. L., & Kutz, J. N. (2019). *Data-Driven Science and Engineering* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108380690>
- Buchmüller, W., & Tye, S. H. H. (1980). Vibrational states in the spectroscopy. *Physical Review Letters*, 44(13), 850–853. <https://doi.org/10.1103/PhysRevLett.44.850>
- Buddenbaum, H., & Steffens, M. (2012). *The Effects of Spectral Pretreatments on Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy*. 2012. <https://doi.org/10.1155/2012/274903>
- Burger, J., & Geladi, P. (2007). Spectral pre-treatments of hyperspectral near infrared images: Analysis of diffuse reflectance scattering. *Journal of Near Infrared Spectroscopy*, 15(1), 29–37.
<https://doi.org/10.1255/jnirs.717>
- Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J., & Smaling, E. M. A. (2012). Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. *Geoderma*, 183–184, 41–48. <https://doi.org/10.1016/j.geoderma.2012.03.011>
- Cameron, E. K., Martins, I. S., Lavelle, P., Mathieu, J., Tedersoo, L., Gottschall, F., Guerra, C. A., Hines, J., Patoine, G., Siebert, J., Winter, M., Cesarz, S., Delgado-Baquerizo, M., Ferlian, O., Fierer, N., Kreft, H., Lovejoy, T. E., Montanarella, L., Orgiazzi, A., ... Eisenhauer, N. (2018). Global gaps in soil biodiversity data. *Nature Ecology and Evolution*, 2(7), 1042–1043. <https://doi.org/10.1038/s41559-018-0573-8>
- Cao, Z., Kühn, P., He, J.-S., Bauhus, J., Guan, Z.-H., & Scholten, T. (2022). Calibration of Near-Infrared Spectra for Phosphorus Fractions in Grassland Soils on the Tibetan Plateau. *Agronomy*, 12(4), 783.
<https://doi.org/10.3390/agronomy12040783>
- Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1), 103–112.
<https://doi.org/10.1016/j.chemolab.2004.12.011>
- Ciani, A., Goss, K. U., & Schwarzenbach, R. P. (2005). Light penetration in soil and particulate minerals. *European Journal of Soil Science*, 56(5), 561–574. <https://doi.org/10.1111/j.1365-2389.2005.00688.x>
- Clark, R. N., King, T. V. V., Klejwa, M., Swayze, G. A., & Vergo, N. (1990). High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research*, 95(B8).

<https://doi.org/10.1029/jb095ib08p12653>

- Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, 89(B7), 6329–6340. <https://doi.org/10.1029/JB089iB07p06329>
- Cootes, T. F., Ionita, M. C., Lindner, C., & Sauer, P. (2012). Robust and accurate shape model fitting using random forest regression voting. In A. Fitzgibbon (Ed.), *Lecture Notes in Computer Science: Vol. 7578 LNCS* (Issue PART 7, pp. 278–291). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33786-4_21
- Coutinho, M. A. N., Alari, F. de O., Ferreira, M. M. C., & Amaral, L. R. d. (2019). Influence of soil sample preparation on the quantification of NPK content via spectroscopy. *Geoderma*, 338(July 2018), 401–409. <https://doi.org/10.1016/j.geoderma.2018.12.021>
- D. C. Slaughter, M. G. Pelletier, & S. K. Upadhyaya. (2001). Sensing Soil Moisture Using Nir Spectroscopy. *Applied Engineering in Agriculture*, 17(2), 241–247. <https://doi.org/10.13031/2013.5449>
- D'Acqui, L. P., Pucci, A., & Janik, L. J. (2010). Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. *European Journal of Soil Science*, 61(6), 865–876. <https://doi.org/10.1111/j.1365-2389.2010.01301.x>
- Dhaliwal, S. S., Naresh, R. K., Mandal, A., Walia, M. K., Gupta, R. K., Singh, R., & Dhaliwal, M. K. (2019). Effect of manures and fertilizers on soil physical properties, build-up of macro and micronutrients and uptake in soil under different cropping systems: a review. *Journal of Plant Nutrition*, 42(20), 2873–2900. <https://doi.org/10.1080/01904167.2019.1659337>
- Ding, J., Yang, A., Wang, J., Sagan, V., & Yu, D. (2018). Machine-learning-based quantitative estimation of soil organic carbon content by VIS/NIR spectroscopy. *PeerJ*, 2018(10), 1–24. <https://doi.org/10.7717/peerj.5714>
- Domingos, P. (2015). *The Master Algorithm*. Perseus Books Group.
- Donald, P., Gary, L., Kriz, G., & Vyvyan, J. (2017). Introduction to Spectroscopy. In *Spectroscopic Methods in Food Analysis* (5th ed.). Cengage Learning. <https://doi.org/10.1201/b21879-2>
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241–258. <https://doi.org/10.1007/s11704-019-8208-z>
- Dotto, A. C., Dalmolin, R. S. D., Grunwald, S., ten Caten, A., & Pereira Filho, W. (2017). Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. *Soil and Tillage Research*, 172(December 2015), 59–68. <https://doi.org/10.1016/j.still.2017.05.008>
- Ehrenfeld, J. G., Ravit, B., & Elgersma, K. (2005). Feedback in the plant-soil system. *Annual Review of Environment and Resources*, 30, 75–115. <https://doi.org/10.1146/annurev.energy.30.050504.144212>
- Eiben, A. E. (2015). Introduction to Evolutionary Computing. In *Kybernetes* (Vol. 33). <https://doi.org/10.1108/03684920410699216>
- El-Zeiny, M. B., Zawbaa, H. M., & Serag, A. (2021). An evaluation of different bio-inspired feature selection techniques on multivariate calibration models in spectroscopy. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 246, 119042. <https://doi.org/10.1016/j.saa.2020.119042>
- Fei, Q., Li, M., Wang, B., Huan, Y., Feng, G., & Ren, Y. (2009). Analysis of cefalexin with NIR spectrometry coupled to artificial neural networks with modified genetic algorithm for wavelength selection. *Chemometrics and Intelligent Laboratory Systems*, 97(2), 127–131. <https://doi.org/10.1016/j.chemolab.2009.03.003>
- Gholizadeh, A., Saberioon, M., Ben-Dor, E., Viscarra Rossel, R. A., & Borůvka, L. (2020). Modelling potentially toxic elements in forest soils with vis-NIR spectra and learning algorithms. *Environmental Pollution*, 267. <https://doi.org/10.1016/j.envpol.2020.115574>
- Gobrecht, A., Roger, J. M., & Bellon-Maurel, V. (2014). Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation. A Review. In *Advances in Agronomy* (Vol. 123). <https://doi.org/10.1016/B978-0-12-420225-2.00004-2>

- Gondal, A. H., Hussain, I., Ijaz, A. bakar, Zafar, A., Ch, B. I., Zafar, H., Sohail, M. D., Niazi, H., Touseed, M., Khan, A. A., Tariq, M., Yousuf, H., & Usama, M. (2021). Influence of soil pH and microbes on mineral solubility and plant nutrition: A review . *International Journal of Agriculture and Biological Sciences*, 5(1), 2–12. https://d1wqtxts1xzle7.cloudfront.net/66113631/Volume_5_Issue_1_Paper_8-with-cover-page-v2.pdf?Expires=1639806217&Signature=FRqt9IzQuEtA1LG0eYffRC1830B80os5kz3-tDqJEqNJJWAdYnhZ8L00~uKz-5hrWLUZmCUQgimABd9LjVhOc1X44eq-W-rC8HtbTBEQeJWWTwNzo0D97Di4NnH1DQcBNJY
- Gonzalez, A., & Vargas, N. (2014). Análisis de imágenes hiperespectrales To cite this version : HAL Id : hal-00935014. *Hal*, 9, 14–17. <https://hal.archives-ouvertes.fr/hal-00935014/document>
- González, M., & Montaña, L. (2015). La espectroscopia y su tecnología: Un repaso histórico y su importancia para el siglo XXI. *Latin-American Journal of Physics Education*, 9(4), 13.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144. <https://doi.org/10.1145/3422622>
- Guiñón, J. L., Ortega, E., García-Antón, J., & Pérez-herranz, V. (2007). Moving Average and Savitzki-Golay Smoothing Filters Using Mathcad. *International Conference on Engineering Education*, 1, 1–4. <http://academic.research.microsoft.com/Paper/12119855.aspx>
- Guo, P., Li, T., Gao, H., Chen, X., Cui, Y., & Huang, Y. (2021). Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using vis-nir spectroscopy. *Remote Sensing*, 13(19), 1–20. <https://doi.org/10.3390/rs13194000>
- Hastie, T., Tibshirani, R., & Friedman, J. (2005). The Elements of Statistical Learning : Data Mining , Inference and Prediction Probability Theory : The Logic of Science The Fundamentals of Risk Measurement Mathematicians , pure and applied , think there is something weirdly different about. *The Mathematical Intelligencer*, 27(2), 83–85. <http://link.springer.com/article/10.1007/BF02985802?LI=true#>
- He, Y., Huang, M., García, A., Hernández, A., & Song, H. (2007). Prediction of soil macronutrients content using near-infrared spectroscopy. *Computers and Electronics in Agriculture*, 58(2), 144–153. <https://doi.org/10.1016/j.compag.2007.03.011>
- He, Y., Xiao, S., Nie, P., Dong, T., Qu, F., & Lin, L. (2017). Research on the Optimum Water Content of Detecting Soil Nitrogen Using Near Infrared Sensor. *Sensors (Switzerland)*, 17(2045), 1–12. <https://doi.org/10.3390/s17092045>
- Helland, I. S. (1990). Partial Least Squares Regression and Statistical Models. *Scandinavian Journal of Statistics*, 17(2), 97–114. <http://www.jstor.org/stable/4616159>
- Hollas, J. M. (2011). Modern Spectroscopy. In *Voenna-meditsinskiĭ zhurnal* (Vol. 332, Issue 2).
- Hong, Y., Chen, Y., Yu, L., Liu, Y., Liu, Y., Zhang, Y., Liu, Y., & Cheng, H. (2018). Combining fractional order derivative and spectral variable selection for organic matter estimation of homogeneous soil samples by VIS-NIR spectroscopy. *Remote Sensing*, 10(3). <https://doi.org/10.3390/rs10030479>
- Horta, M. D. C., & Torrent, J. (2007). The Olsen P method as an agronomic and environmental test for predicting phosphate release from acid soils. *Nutrient Cycling in Agroecosystems*, 77(3), 283–292. <https://doi.org/10.1007/s10705-006-9066-2>
- Huang, G., Zhu, Q., & Siew, C. (2006). Extreme learning machine : Theory and applications. *Neurocomputing*, 70, 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
- Inoue, Y. (2020a). Satellite- and drone-based remote sensing of crops and soils for smart farming—a review. *Soil Science and Plant Nutrition*, 66(6), 798–810. <https://doi.org/10.1080/00380768.2020.1738899>
- Inoue, Y. (2020b). Soil Science and Plant Nutrition Satellite- and drone-based remote sensing of crops and soils for smart farming – a review. *Soil Science and Plant Nutrition*, 66(6), 798–810. <https://doi.org/10.1080/00380768.2020.1738899>
- Isaac, W., & Na, A. (2017). On-the-go soil nitrogen sensor based on near infrared spectroscopy. 2016

- Isaak, S., Yusof, Y., Ngajikin, N. H., Ramli, N., & Wen, C. M. (2019). A low cost spectroscopy with Raspberry Pi for soil macronutrient monitoring. *Telkomnika (Telecommunication Computing Electronics and Control)*, 17(4), 1867–1875. <https://doi.org/10.12928/TELKOMNIKA.V17I4.12775>
- Ji-yong, S., Xiao-bo, Z., Xiao-wei, H., Jie-wen, Z., Yanxiao, L., Limin, H., & Jianchun, Z. (2013). *Rapid detecting total acid content and classifying different types of vinegar based on near infrared spectroscopy and least-squares support vector machine*. 138, 192–199.
<https://doi.org/10.1016/j.foodchem.2012.10.060>
- Ji, W., Adamchuk, V. I., Biswas, A., Dhawale, N. M., Sudarsan, B., Zhang, Y., Viscarra Rossel, R. A., & Shi, Z. (2016). Assessment of soil properties in situ using a prototype portable MIR spectrometer in two agricultural fields. *Biosystems Engineering*, 152, 14–27.
<https://doi.org/10.1016/j.biosystemseng.2016.06.005>
- Ji, W., Li, S., Chen, S., Shi, Z., Viscarra Rossel, R. A., & Mouazen, A. M. (2016). Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. *Soil and Tillage Research*, 155, 492–500. <https://doi.org/10.1016/j.still.2015.06.004>
- Jiang, C., Zhao, J., Ding, Y., & Li, G. (2023). Vis-NIR Spectroscopy Combined with GAN Data Augmentation for Predicting Soil Nutrients in Degraded Alpine Meadows on the Qinghai-Tibet Plateau. *Sensors (Basel, Switzerland)*, 23(7). <https://doi.org/10.3390/s23073686>
- Jin, J., Chen, Z., Li, L., Steponavicius, R., Thennadil, S. N., Yang, J., & Yu, R. (2012). Quantitative Spectroscopic Analysis of Heterogeneous Mixtures : The Properties of Samples. *Analytical Chemistry*, 320–326.
- Kang, H., Gao, H., & Yu, W. (2017). Evaluation of Spectral Pretreatments , Spectral Range and Regression Methods for Quantitative Spectroscopic Analysis of Soil Organic Carbon Composition. *Spectroscopy Letters*, 7010(March). <https://doi.org/10.1080/00387010.2017.1297956>
- Kawamura, K., Nishigaki, T., Tsujimoto, Y., Andriamananjara, A., Rabenaribo, M., Asai, H., Rakotoson, T., & Razafimbelo, T. (2021). Exploring relevant wavelength regions for estimating soil total carbon contents of rice fields in Madagascar from Vis-NIR spectra with sequential application of backward interval PLS. *Plant Production Science*, 24(1), 1–14. <https://doi.org/10.1080/1343943X.2020.1785898>
- Kawamura, K., Tsujimoto, Y., Rabenarivo, M., Asai, H., Andriamananjara, A., & Rakotoson, T. (2017). Vis-NIR spectroscopy and PLS regression with waveband selection for estimating the total C and N of paddy soils in Madagascar. *Remote Sensing*, 9(10). <https://doi.org/10.3390/rs9101081>
- Keyvan, K., Sohrabi, M. R., & Motiee, F. (2021). An intelligent method based on feed-forward artificial neural network and least square support vector machine for the simultaneous spectrophotometric estimation of anti hepatitis C virus drugs in pharmaceutical formulation and biological fluid. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 263, 120190.
<https://doi.org/10.1016/j.saa.2021.120190>
- Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., & Abbas, A. (2018). Modern Trends in Hyperspectral Image Analysis: A Review. *IEEE Access*, 6(c), 14118–14129.
<https://doi.org/10.1109/ACCESS.2018.2812999>
- Kim, H. J., Sudduth, K. A., & Hummel, J. W. (2009). Soil macronutrient sensing for precision agriculture. *Journal of Environmental Monitoring*, 11(10), 1810–1824. <https://doi.org/10.1039/b906634a>
- Kose, U., Prasath, V. B. S., Mondal, M. R. H., Podder, P., & Bharati, S. (2022). Artificial Intelligence and Smart Agriculture Applications. In *Artificial Intelligence and Smart Agriculture Applications*.
<https://doi.org/10.1201/9781003311782>
- Laskar, S., & Mukherjee, S. (2016). Optical Sensing Methods for Assessment of Soil Macro-nutrients and other Properties for Application in Precision Agriculture: A review. *ADB U-Journal of Engineering Technology AJET*, 4(1), 206.
- Legnaioli, S., Lorenzetti, G., Pardini, L., Cavalcanti, G. H., & Palleschi, V. (2014). Double and multiple pulse LIBS techniques. In *Laser-Induced Breakdown Spectroscopy* (Vol. 182, pp. 117–141).

https://doi.org/10.1007/978-3-642-45085-3_5

- Lei, T., & Sun, D. W. (2022). Achieving joint calibration of soil Vis-NIR spectra across instruments, soil types and properties by an attention-based spectra encoding-spectra/property decoding architecture. *Geoderma*, 405(May 2021), 115449. <https://doi.org/10.1016/j.geoderma.2021.115449>
- Lei, Z., Yao, M., Liu, M., Li, Q., & Mao, H. (2011). Comparison between fertilization N, P, K and No fertilization N, P, K in paddy soil by laser induced breakdown spectroscopy. *Proceedings - 4th International Conference on Intelligent Computation Technology and Automation, ICICTA 2011*, 1, 363–366. <https://doi.org/10.1109/ICICTA.2011.102>
- Li, H., Jia, S., & Le, Z. (2019). Quantitative Analysis of Soil Total Nitrogen Using Hyperspectral Imaging Technology with Extreme Learning Machine. *Sensors (Switzerland)*.
- Li, Hong, Wang, J., Zhang, J., Liu, T., Acquah, G. E., & Yuan, H. (2022). Combining Variable Selection and Multiple Linear Regression for Soil Organic Matter and Total Nitrogen Estimation by DRIFT-MIR Spectroscopy. *Agronomy*, 12(3). <https://doi.org/10.3390/agronomy12030638>
- Li, Y., Yang, Q., Chen, M., Wang, M., & Zhang, M. (2019). An ISE-based On-Site Soil Nitrate Nitrogen Detection System. *Sensors (Switzerland)*, 3.
- Liakos, K. G., Busato, P., Moshou, D., & Pearson, S. (2018). Machine Learning in Agriculture : A Review. *Sensors (Basel, Switzerland)*, 18, 1–29. <https://doi.org/10.3390/s18082674>
- Lin, Z. D., Wang, Y. B., Wang, R. J., Wang, L. S., Lu, C. P., Zhang, Z. Y., Song, L. T., & Liu, Y. (2017). Improvements of the Vis-NIRS Model in the Prediction of Soil Organic Matter Content Using Spectral Pretreatments, Sample Selection, and Wavelength Optimization. *Journal of Applied Spectroscopy*, 84(3), 529–534. <https://doi.org/10.1007/s10812-017-0505-4>
- Linker, R., Shmulevich, I., Kenny, A., & Shaviv, A. (2005). Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy. *Chemosphere*, 61(5), 652–658. <https://doi.org/10.1016/j.chemosphere.2005.03.034>
- Liu, J., Xie, J., Han, J., Wang, H., Sun, J., Li, R., & Li, S. (2020). Visible and near-infrared spectroscopy with chemometrics are able to predict soil physical and chemical properties. *Journal of Soils and Sediments*, 20(7), 2749–2760. <https://doi.org/10.1007/s11368-020-02623-1>
- Liu, Lanfa, Ji, M., & Buchroithner, M. (2018). Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. *Sensors (Switzerland)*, 18(9). <https://doi.org/10.3390/s18093169>
- Liu, Lijuan, Shen, B., & Wang, X. (2014). Research on kernel function of support vector machine. *Lecture Notes in Electrical Engineering*, 260 LNEE, 827–834. https://doi.org/10.1007/978-94-007-7262-5_93
- Lourembam, D., Laskar, S., & Mukherjee, S. (2017). Framework for an optical sensor system for monitoring of soil nitrogen and tailoring soil pH. *Journal of Optics*. <https://doi.org/10.1007/s12596-017-0434-x>
- Lv, N., Ma, H., Member, S., Chen, C., & Member, S. (2021). Remote Sensing Data Augmentation Through Adversarial Training. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 9318–9333. <https://doi.org/10.1109/JSTARS.2021.3110842>
- Lyon, L. A., Keating, C. D., Fox, A. P., Baker, B. E., He, L., Nicewarner, S. R., Mulvaney, S. P., & Natan, M. J. (1998). Raman Spectroscopy. *Analytical Chemistry*, 70(12), 341–361.
- Ma, F., Du, C. W., Zhou, J. M., & Shen, Y. Z. (2019). Investigation of soil properties using different techniques of mid-infrared spectroscopy. *European Journal of Soil Science*, 70(1), 96–106. <https://doi.org/10.1111/ejss.12741>
- Ma, J., Cheng, J., Wang, J., Pan, R., He, F., Yan, L., & Xiao, J. (2022). Rapid detection of total nitrogen content in soil based on hyperspectral technology. *Information Processing in Agriculture*, 9(4), 566–574. <https://doi.org/10.1016/j.inpa.2021.06.005>
- MacAbiog, R. E. N., Fadchar, N. A., & Cruz, J. C. D. (2020). Soil NPK Levels Characterization Using Near Infrared and Artificial Neural Network. *Proceedings - 2020 16th IEEE International Colloquium on Signal Processing and Its Applications, CSPA 2020*, Cspa, 141–145. <https://doi.org/10.1109/CSPA48992.2020.9068717>

- Mammadov, E., Denk, M., Riedel, F., Ka, C., Lewinska, K., Łukowiak, R., Grzebisz, W., & Mamedov, A. I. (2022). *Determination of Mehlich 3 Extractable Elements with Visible and Near Infrared Spectroscopy in a Mountainous Agricultural Land , the Caucasus Mountains*.
- Manolakis, D., Lockwood, R., & Cooley, T. (2016). Hyperspectral Imaging Remote Sensing_Physics, Sensors, and Algorithms. In *Hyperspectral Imaging Remote Sensing*.
- Masrie, M., Rosli, A. Z. M., Sam, R., Janin, Z., & Nordin, M. K. (2019). Integrated optical sensor for NPK Nutrient of Soil detection. *2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application, ICSIMA 2018, November, 28–30*.
<https://doi.org/10.1109/ICSIMA.2018.8688794>
- Masrie, M., Rosman, M. S. A., Sam, R., & Janin, Z. (2018). Detection of nitrogen, phosphorus, and potassium (NPK) nutrients of soil using optical transducer. *2017 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2017, 2017-Novem(November), 1–4*.
<https://doi.org/10.1109/ICSIMA.2017.8312001>
- Mendes, W. de S., Sommer, M., Koszinski, S., & Wehrhan, M. (2021). Local Peatlands Spectral Data Influence in Global Spectral Modelling of Soil Organic Carbon and Total Nitrogen Using Near-Infrared Spectrum. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3983929>
- Min, M., & Lee, W. S. (2005). Determination of Significant Wavelengths and Prediction of Nitrogen Content for Citrus. *American Society of Agricultural Engineers ISSN, 48(1998), 455–461*.
- Mirzaee, S., Ghorbani-Dashtaki, S., Mohammadi, J., Asadi, H., & Asadzadeh, F. (2016). Spatial variability of soil organic matter using remote sensing data. *Catena, 145, 118–127*.
<https://doi.org/10.1016/j.catena.2016.05.023>
- Mohamed, E. S., El Baroudy, A. A., El-beshbeshy, T., Emam, M., Belal, A. A., Elfadaly, A., Aldosari, A. A., Ali, A. M., & Lasaponara, R. (2020). Vis-nir spectroscopy and satellite landsat-8 oli data to map soil nutrients in arid conditions: A case study of the northwest coast of egypt. *Remote Sensing, 12(22), 1–20*. <https://doi.org/10.3390/rs12223716>
- Montesinos-López, O. A., Montesinos-López, A., & Corssa, J. (2022). Over fitting , Model Tuning , and Evaluation of Prediction Performance. In F. van Euwijk (Ed.), *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 109–139). Springer.
- Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Wiebensohn, J., Bill, R., & Mouazen, A. M. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering, 152, 104–116*.
<https://doi.org/10.1016/j.biosystemseng.2016.04.018>
- Moros Portolés, J. (2007). *TRATAMIENTO NUMÉRICO DE LOS DATOS EN EL ANÁLISIS CUANTITATIVO POR ESPECTROMETRÍA VIBRACIONAL* (Vol. 3, Issue September). Universitat de Valencia.
- Mukherjee, S., & Laskar, S. (2019). Vis–NIR-based optical sensor system for estimation of primary nutrients in soil. *Journal of Optics (India), 48(1), 87–103*. <https://doi.org/10.1007/s12596-019-00517-1>
- Munnaf, M. A., Guerrero, A., Nawar, S., Haesaert, G., Van Meirvenne, M., & Mouazen, A. M. (2021). A combined data mining approach for on-line prediction of key soil quality indicators by Vis-NIR spectroscopy. *Soil and Tillage Research, 205(June 2020), 104808*.
<https://doi.org/10.1016/j.still.2020.104808>
- Murphy, K. P. (2012). Machine Learning A Probabilistic Perspective. In *Chance Encounters: Probability in Education*. https://doi.org/10.1007/978-94-011-3532-0_2
- Nawar, S., Delbecq, N., Declercq, Y., De Smedt, P., Finke, P., Verdoodt, A., Van Meirvenne, M., & Mouazen, A. M. (2019). Can spectral analyses improve measurement of key soil fertility parameters with X-ray fluorescence spectrometry? *Geoderma, 350(May), 29–39*.
<https://doi.org/10.1016/j.geoderma.2019.05.002>
- Nie, P. C., Dong, T., He, Y., & Qu, F. (2017). Detection of soil nitrogen using near infrared sensors based on soil pretreatment and algorithms. *Sensors (Switzerland), 17(5), 1–13*.
<https://doi.org/10.3390/s17051102>

- Nie, P., Dong, T., He, Y., & Xiao, S. (2018). Research on the Effects of Drying Temperature on Nitrogen Detection of Different Soil Types by Near. *Sensors (Switzerland)*. <https://doi.org/10.3390/s18020391>
- Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B., & Montanarella, L. (2014). Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biology and Biochemistry*, 68, 337–347. <https://doi.org/10.1016/j.soilbio.2013.10.022>
- Norris, K. H., & Williams, P. C. (1984). Optimization of Mathematical Treatments of Raw Near-Infrared Signal in the Measurement of Protein in Hard Red Spring Wheat. *Cereal Chemistry*, 2, 158–165.
- Papousek, D., & Aliev, M. R. (1982). *Molecular Vibration-Rotation Spectra* (R. Zahradnik (ed.); 1st ed.). Elsevier.
- Patel, A. K., Ghosh, J. K., & Sayyad, S. U. (2020). Fractional abundances study of macronutrients in soil using hyperspectral remote sensing. *Geocarto International*, 0(0), 1–20. <https://doi.org/10.1080/10106049.2020.1720315>
- Petropoulos, G. P., Arvanitis, K., & Sigrimis, N. (2012). Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Expert Systems with Applications*, 39(3), 3800–3809. <https://doi.org/10.1016/j.eswa.2011.09.083>
- Qi, H., Paz-Kagan, T., Karnieli, A., Jin, X., & Li, S. (2018). Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. *Soil and Tillage Research*, 175(October 2017), 267–275. <https://doi.org/10.1016/j.still.2017.09.006>
- Reeves, J. B. (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*, 158(1–2), 3–14. <https://doi.org/10.1016/j.geoderma.2009.04.005>
- 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-Pérez, J. R., Marcelo, V., Pereira-Obaya, D., García-Fernández, M., & Sanz-Ablanedo, E. (2021). Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards. *Agronomy*, 11(10). <https://doi.org/10.3390/agronomy11101895>
- Rojas, R. (1996). The Backpropagation Algorithm. In *Neural Networks* (pp. 150–182). Springer Berlin Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-61068-4_7
- Sadgrove, E. J., Falzon, G., Miron, D., & Lamb, D. W. (2021). The segmented colour feature extreme learning machine: Applications in agricultural robotics. *Agronomy*, 11(11), 1–16. <https://doi.org/10.3390/agronomy11112290>
- Sáez-Plaza, P., García, A., & Martín, J. (2019). An annotation on the Kjeldahl method. *Anales de La Real Academia Nacional de Farmacia*, 85(1), 14–19.
- Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2018). Recent Advances in Recurrent Neural Networks. *Neural and Evolutionary Computing*, 1–21.
- Sarathjith, M. C., Das, B. S., Wani, S. P., & Sahrawat, K. L. (2016). Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma*, 267, 1–9. <https://doi.org/10.1016/j.geoderma.2015.12.031>
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627–1639.
- Norma Oficial Mexicana NOM-021-RECNAT-2000, Diario Oficial de la Federación (2002). http://dof.gob.mx/nota_detalle.php?codigo=717582&fecha=31/12/2002
- Shao, Y., & He, Y. (2011). Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy. *Soil Research*, 49(2), 166–172. <https://doi.org/10.1071/SR10098>
- Shepherd, K. D., & Walsh, M. G. (2002). Development of Reflectance Spectral Libraries for

- Characterization of Soil Properties. *Soil Science Society of America Journal*, 66(3), 988–998.
<https://doi.org/10.2136/sssaj2002.9880>
- Shi, P., Castaldi, F., van Wesemael, B., & Van Oost, K. (2020). Vis-NIR spectroscopic assessment of soil aggregate stability and aggregate size distribution in the Belgian Loam Belt. *Geoderma*, 357(May 2019), 113958. <https://doi.org/10.1016/j.geoderma.2019.113958>
- Singh, R. (2002). C. V. Raman and the Discovery of the Raman Effect. *Physics in Perspective*, 4(4), 399–420. <https://doi.org/10.1007/s000160200002>
- Sleep, B., Mason, S., Janik, L., & Mosley, L. (2022). Application of visible near-infrared absorbance spectroscopy for the determination of Soil pH and liming requirements for broad-acre agriculture. *Precision Agriculture*, 23(1), 194–218. <https://doi.org/10.1007/s11119-021-09834-7>
- Song, X., Gao, Y., Liu, Z., Zhang, M., & Wan, Y. (2019). Development of a predictive tool for rapid assessment of soil total nitrogen in wheat-corn double cropping system with hyperspectral data. *Environmental Pollutants and Bioavailability*, 5940. <https://doi.org/10.1080/26395940.2019.1679041>
- 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>
- Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE*, 8(6). <https://doi.org/10.1371/journal.pone.0066409>
- Stevens, A., & Ramirez-Lopez, L. (2013). An Introduction to the prospectr package. In *2013 R Package* (0.1; Vol. 136, Issue 12, pp. 1628-a-1629). <https://doi.org/10.1176/ajp.136.12.1628-a>
- Sun, J., Shi, S., Yang, J., Gong, W., Qiu, F., Wang, L., Du, L., & Chen, B. (2019). Wavelength selection of the multispectral lidar system for estimating leaf chlorophyll and water contents through the PROSPECT model. *Agricultural and Forest Meteorology*, 266–267(November 2018), 43–52. <https://doi.org/10.1016/j.agrformet.2018.11.035>
- Sutton, R. S., & Barto, A. G. (2019). Reinforcement Learning, An introduction. In MIT (Ed.), 余东华 张鑫宇 孙婷 (2nd ed.). Westchester Publishing Services.
- Thenkabail, P. S., Lyon, J. G., & Huete, A. (2018). HYPERSPECTRAL REMOTE SENSING OF VEGETATION. In P. S. Thenkabail, G. J. Lyon, & A. Huete (Eds.), *Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation* (2nd ed.). CRC Press. <https://doi.org/10.1201/9781315164151>
- Thurnhofer-hemsi, K., López-rubio, E., & Molina-cabello, M. A. (2020). RADIAL BASIS FUNCTION KERNEL OPTIMIZATION FOR SUPPORT VECTOR MACHINE CLASSIFIERS. *IEEE Transactions on Neural Networks and Learning Systems*. <https://doi.org/https://doi.org/10.48550/arXiv.2007.08233>
- Tümsavaş, Z. (2017). Application of visible and near infrared reflectance spectroscopy to predict total nitrogen in soil. *Environmental Biology*, 38(September), 409–417. [https://doi.org/10.22438/jeb/38/5\(SI\)/GM-047](https://doi.org/10.22438/jeb/38/5(SI)/GM-047)
- Tziolas, N., Tsakiridis, N., Ogen, Y., Kalopesa, E., Ben-Dor, E., Theocharis, J., & Zalidis, G. (2020). An integrated methodology using open soil spectral libraries and Earth Observation data for soil organic carbon estimations in support of soil-related SDGs. *Remote Sensing of Environment*, 244(January), 111793. <https://doi.org/10.1016/j.rse.2020.111793>
- Uhlenbeck, G. E., & Goudsmit, S. (1976). Spinning Electrons and the Structure of Spectra. *Nature*, 2(4), 391. [https://doi.org/10.1016/0304-3762\(76\)90072-9](https://doi.org/10.1016/0304-3762(76)90072-9)
- United States Department of Agriculture. (1987). USDA Textural Soil Classification. In *Soil Mechanics Level I Module 3 - USDA Textural Soil Classification* (pp. 1–53).
- Van Reeuwijk, L. P. (2002). *Procedures For Soil Analysis* (L. P. Van Reeuwijk (ed.); 6th ed.). International Soil Reference and Information Centre.

- Vanderlinde, J., & Neuenschwander, D. E. (1994). Classical Electromagnetic Theory. In *American Journal of Physics* (Vol. 62, Issue 7). <https://doi.org/10.1119/1.17492>
- Vapnik, V. (1998). *of Function Estimation*. 55–85. https://doi.org/10.1007/978-1-4615-5703-6_
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5), 988–999. <https://doi.org/10.1109/72.788640>
- Vestergaard, R.-J., Adamchuk, V., & Biswas, A. (2021). Evaluation of Optimized Preprocessing and Modeling Algorithms for Prediction of Soil Properties Using VIS-NIR Spectroscopy. *Sensors (Basel, Switzerland)*. <https://doi.org/doi.org/10.3390/s21206745>
- Vestergaard, R., Vasava, H., Aspinall, D., Chen, S., Gillespie, A., Adamchuk, V., & Biswas, A. (2021). *Evaluation of Optimized Preprocessing and Modeling*. 3.
- Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., ... Ji, W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155, 198–230. <https://doi.org/10.1016/j.earscirev.2016.01.012>
- Viscarra Rossel, R. A., Rizzo, R., Demattê, J. A. M., & Behrens, T. (2010). Spatial Modeling of a Soil Fertility Index using Visible-Near-Infrared Spectra and Terrain Attributes. *Soil Science Society of America Journal*, 74(4), 1293–1300. <https://doi.org/10.2136/sssaj2009.0130>
- 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(1), 88–96. <https://doi.org/10.1016/j.geoderma.2014.01.013>
- Wambugu, N., Chen, Y., Xiao, Z., Tan, K., Wei, M., Liu, X., & Li, J. (2021). Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks : A review. *International Journal of Applied Earth Observations and Geoinformation*, 105, 102603. <https://doi.org/10.1016/j.jag.2021.102603>
- Wan, M., Qu, M., Hu, W., Li, W., Zhang, C., Cheng, H., & Huang, B. (2019). Estimation of soil pH using PXRF spectrometry and Vis-NIR spectroscopy for rapid environmental risk assessment of soil heavy metals. *Process Safety and Environmental Protection*, 132, 73–81. <https://doi.org/10.1016/j.psep.2019.09.025>
- Wang, C., Liu, B., Liu, L., Zhu, Y., Hou, J., Liu, P., & Li, X. (2021). A review of deep learning used in the hyperspectral image analysis for agriculture. In *Artificial Intelligence Review* (Vol. 54, Issue 7). Springer Netherlands. <https://doi.org/10.1007/s10462-021-10018-y>
- Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, 324(1). <https://doi.org/10.1088/1757-899X/324/1/012049>
- Warren, G. P., & Whitehead, D. C. (1988). Available soil nitrogen in relation to fractions of soil nitrogen and other soil properties. *Plant and Soil*, 112(2), 155–165. <https://doi.org/10.1007/BF02139991>
- Wartini, N., Minasny, B., Montazerolghaem, M., & Padarian, J. (2019). *Convolutional Neural Network for simultaneous prediction of several soil properties using visible / near-infrared , mid- infrared , and their combined spectra* [University of Sydney]. <https://www.sciencedirect.com/science/article/pii/S0016706119300588>
- Wehrle, R., Welp, G., & Pätzold, S. (2021). Total and hot-water extractable organic carbon and nitrogen in organic soil amendments: Their prediction using portable mid-infrared spectroscopy with support vector machines. *Agronomy*, 11(4). <https://doi.org/10.3390/agronomy11040659>
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236(December 2018), 111402. <https://doi.org/10.1016/j.rse.2019.111402>
- Wenjun, J., Zhou, S., Jingyi, H., & Shuo, L. (2014). In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy. *PLoS ONE*, 9(8). <https://doi.org/10.1371/journal.pone.0105708>

- Wilson, M. J. (2019). The importance of parent material in soil classification: A review in a historical context. *Catena*, 182(May), 104131. <https://doi.org/10.1016/j.catena.2019.104131>
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37–52. https://doi.org/10.1007/978-0-387-87811-9_2
- Wu, M., Wang, S., Pan, S., Terentis, A. C., & Strasswimmer, J. (2021). Deep learning data augmentation for Raman spectroscopy cancer tissue classification. *Scientific Reports*, 1–13. <https://doi.org/10.1038/s41598-021-02687-0>
- Xiao, S., & He, Y. (2019). Application of near-infrared spectroscopy and multiple spectral algorithms to explore the effect of soil particle sizes on soil nitrogen detection. *Molecules*, 24(13). <https://doi.org/10.3390/molecules24132486>
- Xu, G., Li, Z., & Li, P. (2013). Fractal features of soil particle-size distribution and total soil nitrogen distribution in a typical watershed in the source area of the middle Dan River, China. *Catena*, 101, 17–23. <https://doi.org/10.1016/j.catena.2012.09.013>
- Xu, S., Zhao, Y., Wang, M., & Shi, X. (2018). Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. *Geoderma*, 310(July 2017), 29–43. <https://doi.org/10.1016/j.geoderma.2017.09.013>
- Yan, X. T., Donaldson, K. M., Davidson, C. M., & Gao, Y. (2018). Effects of sample pretreatment and particle size on the determination of nitrogen in soil by portable LIBS and potential use on robotic-borne remote Martian and agricultural soil analysis systems. *RSC Advances*, 36886–36894. <https://doi.org/10.1039/C8RA07065B>
- Yang, H., Kuang, B., & Mouazen, A. M. (2012). Quantitative analysis of soil nitrogen and carbon at a farm scale using visible and near infrared spectroscopy coupled with wavelength reduction. *European Journal of Soil Science*, 63(3), 410–420. <https://doi.org/10.1111/j.1365-2389.2012.01443.x>
- Yang, Haiqing. (2011). Spectroscopic calibration for soil N and C measurement at a farm scale. *Procedia Environmental Sciences*, 10(PART A), 672–677. <https://doi.org/10.1016/j.proenv.2011.09.108>
- 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(January), 114616. <https://doi.org/10.1016/j.geoderma.2020.114616>
- Yang, M., Mouazen, A., Zhao, X., & Guo, X. (2020). Assessment of a soil fertility index using visible and near-infrared spectroscopy in the rice paddy region of southern China. *European Journal of Soil Science*, 71(4), 615–626. <https://doi.org/10.1111/ejss.12907>
- Yang, M., Xu, D., Chen, S., Li, H., & Shi, Z. (2019). Evaluation of machine learning approaches to predict soil organic matter and pH using vis-NIR spectra. *Sensors (Switzerland)*, 19(2). <https://doi.org/10.3390/s19020263>
- Yang, X. S. (2009). Firefly algorithms for multimodal optimization. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5792 LNCS, 169–178. https://doi.org/10.1007/978-3-642-04944-6_14
- Yao, Xiangqian, Yang, W., Li, M., Zhou, P., Chen, Y., Hao, Z., & Liu, Z. (2018). Prediction of Total Nitrogen in Soil Based on Random Frog Leaping Wavelet Neural Network. *IFAC-PapersOnLine*, 51(17), 660–665. <https://doi.org/10.1016/j.ifacol.2018.08.121>
- Yao, Xiong, Yu, K., Deng, Y., Liu, J., & Lai, Z. (2019). Spatial variability of soil organic carbon and total nitrogen in the hilly red soil region of Southern China. *Springer*.
- Young, A. T. (1981). Rayleigh scattering. *Applied Optics*, 20(4), 533–535.
- Yuan, J., Wang, X., Yan, C., Chen, S., Wang, S., Zhang, J., Xu, Z., Ju, X., Ding, N., Dong, Y., & Zhang, W. (2020). Wavelength selection for estimating soil organic matter contents through the radiative transfer model. *IEEE Access*, 8, 176286–176293. <https://doi.org/10.1109/ACCESS.2020.3026813>
- Yubing Wang, Huang, H., & Chen, X. (2021). Predicting Organic Matter Content, Total Nitrogen and pH Value of Lime Concretion Black Soil Based on Visible and Near Infrared Spectroscopy. *Eurasian Soil Science*, 54(11), 1681–1688. <https://doi.org/10.1134/S1064229321110144>

- Yun, Y. H., Li, H. D., Deng, B. C., & Cao, D. S. (2019). An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, 113, 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>
- Yun, Y. H., Wang, W. T., Tan, M. L., Liang, Y. Z., Li, H. D., Cao, D. S., Lu, H. M., & Xu, Q. S. (2014). A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Analytica Chimica Acta*, 807, 36–43. <https://doi.org/10.1016/j.aca.2013.11.032>
- Yun, Y., Li, H., Deng, B., & Cao, D. (2019). *Trends in Analytical Chemistry An overview of variable selection methods in multivariate analysis of near-infrared spectra*. 113, 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>
- Zhang, Jian, Gao, B., Chai, H., Ma, Z., & Yang, G. (2016). Identification of DNA-binding proteins using multi-features fusion and binary firefly optimization algorithm. *BMC Bioinformatics*, 17(1), 1–12. <https://doi.org/10.1186/s12859-016-1201-8>
- Zhang, Jifeng, Wang, Z., Fan, B., Hou, Y., Dou, Y., Ren, Z., & Chen, X. (2020). *Investigating the Proper Application Rate of Nitrogen under Mulched Drip Irrigation to Improve the Yield and Quality of Tomato in Saline Soil*.
- Zhang, Y., Li, M. Z., Zheng, L. H., Zhao, Y., & Pei, X. (2016). Soil nitrogen content forecasting based on real-time NIR spectroscopy. *Computers and Electronics in Agriculture*, 124, 29–36. <https://doi.org/10.1016/j.compag.2016.03.016>
- Zhang, Y., Li, M., Zheng, L., Qin, Q., & Suk, W. (2019). Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm. *Geoderma*, 333(June 2018), 23–34. <https://doi.org/10.1016/j.geoderma.2018.07.004>
- Zhou, P., Yang, W., Li, M., Yao, X., & Liu, Z. (2018). Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Tot. *IFAC-PapersOnLine*, 51(17), 51–56. <https://doi.org/10.1016/j.ifacol.2018.08.071>
- Zhou, P., Zhang, Y., Yang, W., Li, M., Liu, Z., & Liu, X. (2019). Development and performance test of an in-situ soil total nitrogen-soil moisture detector based on near-infrared spectroscopy. *Computers and Electronics in Agriculture*, 160(January), 51–58. <https://doi.org/10.1016/j.compag.2019.03.016>

CAPITULO 4. EVALUACIÓN DE ALGORITMOS DE CALIBRACIÓN PARA PREDICCIÓN DE NITRÓGENO DEL SUELO MEDIANTE ESPECTROSCOPÍA NIR Y AUMENTO DE DATOS CON REDES GENERATIVAS ADVERSARIAS.

Resumen

La espectroscopia y el aprendizaje automático desempeñan un papel crucial en la agricultura de precisión, a través de modelos espectrales predictivos es posible mejorar la gestión y manejo de variabilidad de recursos como el suelo. Sin embargo, debido a las relaciones complejas entre los componentes del suelo y su reflectancia, es necesaria la elección de algoritmos de regresión adecuados, a través de una comparación exhaustiva de la precisión alcanzada con los diferentes enfoques de correlación que utiliza cada uno. Asimismo, muchos de estos modelos demandan una gran cantidad de datos para lograr su mejor rendimiento, lo que puede resultar poco asequible. En este estudio, se evaluó la capacidad predictiva de cuatro algoritmos de regresión para la cuantificación de nitrógeno aprovechable del suelo con espectroscopía NIR: la red neuronal de aprendizaje extremo (AE); máquinas de vectores de soporte (MVS); bosques aleatorios (BA); y el modelo lineal de mínimos cuadrados parciales (PLS). El pretratamiento espectral se basó en el promedio móvil y en el filtro derivativo de Savitzky Galay, la selección de variables espectrales se realizó utilizando un Algoritmo Genético (AG). Asimismo, en cada modelo se evaluó el efecto de incrementar el conjunto de entrenamiento, a través de datos espectrales artificiales generados con redes generativas adversarias. Para lo cual, se analizó la calidad de los datos y el impacto en las métricas de evaluación de predicción: raíz del error cuadrado medio, coeficiente de determinación y la Relación de Rendimiento y Desviación (RMSE, R^2 y RPD). El aumento con datos artificiales tuvo un efecto positivo en el rendimiento de los modelos, mejorando las métricas de evaluación. Se observó que MVS y BA tuvieron un mejor desempeño general, en términos de precisión de predicción y aprovechamiento de información, alcanzando RPD superiores a 2, R^2 mayores que 0.8 y menores índices de error en comparación con PLS y AE.

Palabras clave: *detección remota, modelos de regresión, aprendizaje automático, datos artificiales, rendimiento predictivo.*

4.1. INTRODUCCIÓN

En el ámbito de la agricultura de precisión (AP), la detección y manejo de variabilidad del nitrógeno en los suelos, desempeña un papel de vital importancia.

Específicamente los amonios (NH_4^+) y nitratos (NO_3^-) son formas de nitrógeno aprovechable (NA) que puede ser procesado por las plantas. Estos compuestos son absorbidos y aprovechados por la planta para la síntesis de proteínas, ácidos nucleicos y pigmentos (Patel *et al.*, 2020). Además, estos elementos son de los principales constituyentes de los fertilizantes, por lo tanto, la determinación del contenido de NA en suelos, es clave para implementar aplicaciones precisas, optimizando recursos y disminuyendo el impacto ambiental a causa de la aplicación excesiva de insumos.

Una de las técnicas más prometedoras en AP es el uso de imágenes multi e hiperespectrales para detección remota, que capturan información geoespacial y espectral para cada píxel, lo cual, con la metodología adecuada se puede traducir en información útil sobre los nutrientes disponibles en suelos (Khan *et al.*, 2018). Por lo tanto, el empleo de la espectroscopia como herramienta de análisis toma un papel fundamental para AP. Esta es una tecnología no destructiva, con la que se analiza la interacción (absorción, reflectancia o transmitancia) entre la materia y la radiación, proyectada en forma de espectros (Hollas, 2011). Con la espectroscopía se pueden identificar los elementos individuales de una mezcla homogénea, ya que, debido a sus enlaces particulares y estructura molecular, cada especie química produce un espectro de absorción único (Isaac & Na, 2017). Junto con técnicas de aprendizaje automático (AA), se utiliza en ciencias del suelo para analizar características fisicoquímicas y construir modelos espectrales de regresión y clasificación. Numerosos estudios respaldan la utilidad y precisión de estos modelos para determinar muchos atributos del suelo, utilizando específicamente las regiones espectrales del infrarrojo cercano (NIR) y visible (Vis) (Sleep *et al.*, 2022; Viscarra Rossel *et al.*, 2010; Yubing Wang *et al.*, 2021; Zhang *et al.*, 2016, etc.).

Con el avance de los instrumentos analíticos modernos, es posible obtener espectros de mayor resolución, con numerosas variables espectrales (longitudes de onda), lo que brinda información más abundante. Sin embargo, existen varias desventajas, como alta redundancia, alto costo computacional, ruido y otras alteraciones espectrales producto de la dispersión de radiación en el suelo y el

entorno de medición. Por lo que el desarrollo de estos modelos requiere de técnicas de selección de variables y de preprocesamiento de los espectros que reduzcan estos efectos (Kawamura *et al.*, 2021; R. Vestergaard *et al.*, 2021; Xiao & He, 2019; Yuan *et al.*, 2020). Además, los modelos se desarrollan utilizando aprendizaje supervisado, por lo que se requieren métodos de calibración multivariantes para extraer patrones de reflectancia o absorción complejos y correlacionarlos con las propiedades del suelo medidas (Araújo *et al.*, 2014; S. Xu *et al.*, 2018). La regresión de componentes principales (PCR) y la regresión de mínimos cuadrados parciales (PLS) son técnicas que asumen relaciones lineales entre variables y son ampliamente utilizadas en ciencias de suelos (Armenise *et al.*, 2013; Kawamura *et al.*, 2017; M. Yang *et al.*, 2020; Yubing Wang *et al.*, 2021). Sin embargo, recientemente el interés en el uso de técnicas de calibración no lineal está aumentando debido a que las relaciones entre los datos espectrales y las características del suelo rara vez son de naturaleza lineal (Araújo *et al.*, 2014), propiedades como el nitrógeno y otros elementos podrían estar relacionados con la reflectancia tanto de manera lineal como no lineal (Qi *et al.*, 2018). En AA las técnicas no lineales incluyen redes neuronales artificiales (RNA), árboles de decisión (AD) y bosques aleatorios (BA), regresión de máquinas de vectores de soporte (MVS), redes de aprendizaje extremo (AE), modelos cubistas entre otros (Ahmadi *et al.*, 2021; Barra *et al.*, 2021; Liakos *et al.*, 2018; C. Wang *et al.*, 2021). Por lo que, en la literatura científica se puede encontrar una gran cantidad de estudios dedicados a comprender, analizar y comparar algoritmos con diferentes enfoques de regresión que permitan una mejor relación entre la reflectancia y los elementos en el suelo (Guo *et al.*, 2021; Vestergaard *et al.*, 2021; Xu *et al.*, 2018; Yang *et al.*, 2019). Sin embargo, ninguna de estas técnicas de calibración propuestas ha logrado la aceptación universal, ya que, un modelo que funciona para una aplicación en condiciones específicas, puede no ser viable para otros escenarios (S. Xu *et al.*, 2018).

Por otro lado, aunque el AA ha demostrado un gran potencial en estudios relacionados con la teledetección de elementos del suelo, la efectividad de modelos de inteligencia artificial como MVS, BA, AE o PLS está estrechamente

relacionada con la cantidad y la calidad de los datos espectrales utilizados para su entrenamiento (Barra *et al.*, 2021; Cambule *et al.*, 2012; Gholizadeh *et al.*, 2020). Como alternativa, en años recientes modelos generativos basados en aprendizaje profundo como las redes generativas adversarias (GAN por sus siglas en inglés) se han posicionado como una alternativa viable para la escasez de datos. Las GAN son capaces de aprender la distribución de los datos reales y generar ejemplos sintéticos que se asemejen a ellos (Goodfellow *et al.*, 2020). Además, son técnicas que no requiere muchos datos de entrenamiento (Baek *et al.*, 2019) y han demostrado ser viables para reproducir espectros de suelos con el propósito de mejorar el desempeño de modelos AA (Jiang *et al.*, 2023). Este estudio tiene como objetivos; evaluar la capacidad predictiva de los algoritmos PLS, AE, MVS y BA para la estimación cuantitativa del nitrógeno aprovechable en suelos. Además de analizar la calidad y el efecto que tienen los datos espectrales generados con GAN en el entrenamiento y el desempeño predictivo de cada algoritmo de regresión.

4.2. MATERIALES Y METODOS

4.2.1 Área de estudio, muestreo y análisis nutricional

El muestreo para este estudio se realizó en el área experimental “La Xerona”, que corresponde al campo experimental del Departamento de Ingeniería Mecánica Agrícola en la Universidad Autónoma Chapingo. Está ubicado en las coordenadas geográficas 19° 29' 2.3" N de latitud y 98° 53' 59.8" O de longitud y tiene una superficie de 9.31 ha, constituidas por cuatro texturas de suelo predominantes: Franco-limoso, Franco-arcilloso, Franco y Arcilloso (Figura 4.1). El clima predominante en la región es templado subhúmedo, con una estación seca marcada y una precipitación anual moderada.

El suelo fue extraído de la capa superficial (0-30 cm) de cada punto obteniendo un total de 129 muestras. Para su almacenamiento y transporte, se retiraron piedras y residuos vegetales e inorgánicos de las muestras, después fueron secadas en un horno de convección mecánica marca Imperial V modelo 3481M a 90° C por 36 horas. Una vez seco, el suelo fue molido y tamizado con tamiz del

número 10 (2 mm de abertura) para homogeneizar el tamaño de partícula. Cada muestra se dividió por el método de cuarteo en dos porciones, una para análisis nutricional de laboratorio y otra para análisis espectral.

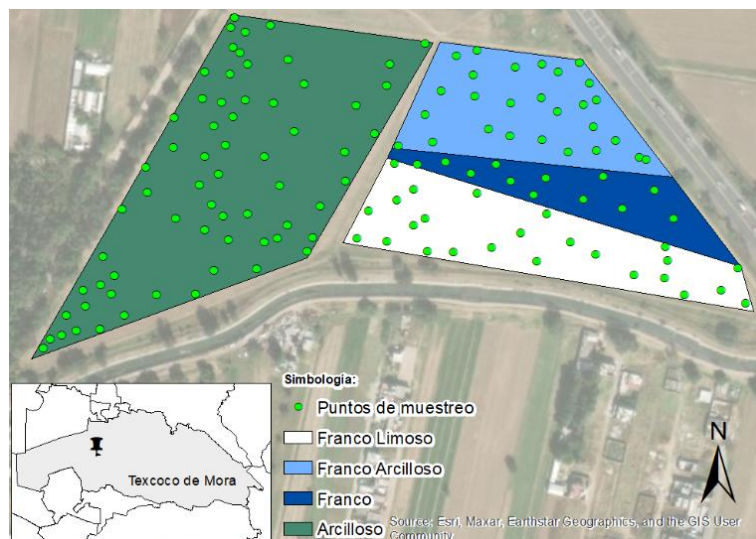


Figura 4.1. Campo Experimental “La Xerona”, Chapingo, Texcoco Estado de México.

Las muestras fueron enviadas al Laboratorio Nacional de Investigación y Servicio Agroalimentario y Forestal (LANISAF) para la determinación de nitrógeno aprovechable para las plantas. Se utilizó el método estándar de Kjeldahl (Sáez-Plaza *et al.*, 2019), el cual es un procedimiento de oxidación húmeda que indica la disponibilidad de nitrógeno (NA) para las plantas en forma de nitratos NO_3^- y amonio NH_4^+ .

4.2.2 Análisis espectral

Los espectros de reflectancia se midieron en LANISAF con un espectrofotómetro de infrarrojo cercano NIR de la marca Bruker® modelo MPA (122000) con un rango de 800 a 2700 nm y resolución espectral de 0.125 nm. Para cada medición, el equipo fue calibrado con un panel Spectralon® con 99% de reflectancia. Para el análisis, cada muestra se colocó en un contenedor cilíndrico de cristal. En virtud de obtener espectros representativos, para cada muestra se realizaron tres repeticiones del análisis y se promediaron.

4.2.3 Desarrollo de los modelos espectrales.

En este estudio, se evaluó la precisión de predicción de NA de suelos con modelos espectrales generados a partir de cuatro algoritmos de regresión. Además, se analizó el efecto del incremento del conjunto de entrenamiento a partir de datos generados artificialmente. Para esto, se utilizaron datos NIR y los datos de NA en un modelo generativo GAN para incrementar el número de datos en el conjunto de entrenamiento. El método técnico se muestra en la Figura 4.2, en primer lugar, los datos espectrales fueron pretratados, después se realizó una división de datos, destinando 95 de ellos para el conjunto inicial de entrenamiento, el resto se designaron para validación. Luego, el conjunto de entrenamiento se utilizó para selección de variables y para entrenar la red GAN y generar espectros y datos de NA artificiales. El conjunto de entrenamiento se utilizó para selección de variables y para entrenar la red GAN y generar espectros y datos de NA artificiales.

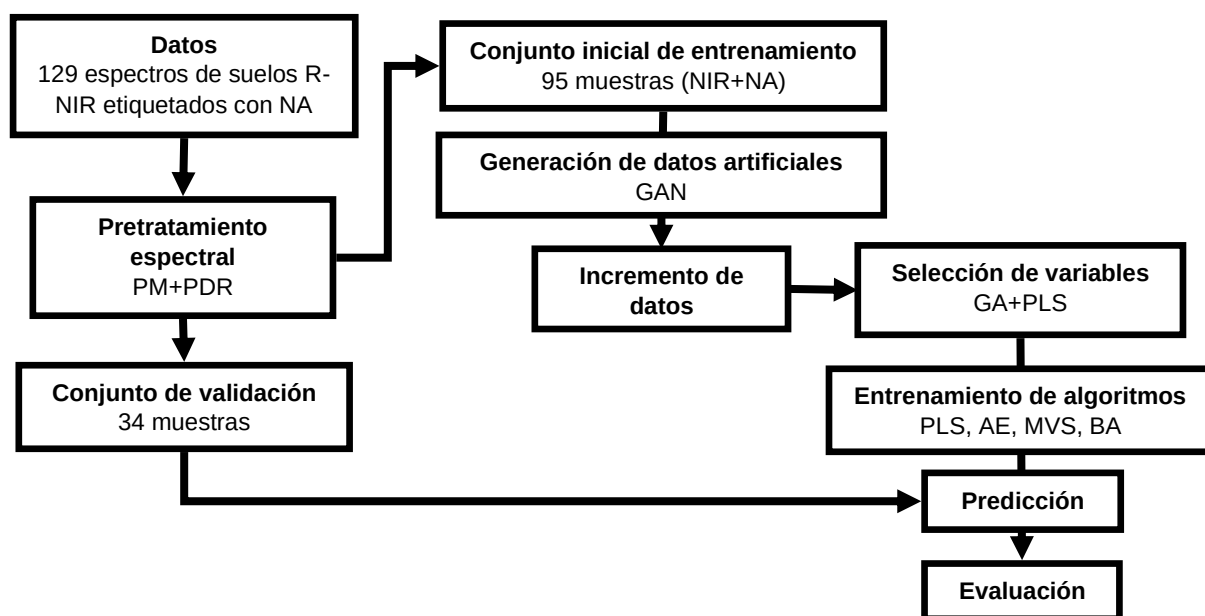


Figura 4.2. Diagrama de los modelos de regresión con aumento de datos

Ambos conjuntos, los de entrenamiento y los generados por GAN se utilizaron para crear conjuntos de entrenamiento de diferentes tamaños y calibrar los algoritmos de regresión PLS, BA, MVS y AE. Finalmente se utilizaron los datos del conjunto de prueba para hacer predicciones con los modelos entrenados en diferentes conjuntos, donde se evaluó el desempeño de cada uno de ellos con las métricas de predicción indicadas en la sección 2.4.1. Los modelos fueron

implementados en Matlab R2020B y las redes GAN en Python 3 en una máquina de 32GB de memoria RAM y procesador Intel Core i7 a 2.90 GHz.

Pretratamiento espectral

Los métodos de pretratamiento de espectro reducen el ruido digital y las interferencias en el espectro por causa de factores no sistemáticos y de dispersión, mejorando así la precisión de los algoritmos predictivos (Kang *et al.*, 2017; Lin *et al.*, 2017). En este estudio se empleó un suavizado espectral de promedio móvil (PM) para la eliminación de ruido de alta frecuencia (Guiñón *et al.*, 2007). PM genera una corrección en la reflectancia de cada variable espectral, promediando la respuesta espectral de un número determinado de longitudes de onda contiguas. La cantidad de variables espectrales determina el tamaño de una ventana móvil, la cual se desplaza sobre todo el espectro original. Así mismo para correcciones de dispersión y resaltar características del espectro se utilizó la primera derivada de la reflectancia (PDR) aplicando el filtro de Savitzky-Golay derivativo (SGD) propuesto por Savitzky & Golay, (1964). Los parámetros de cada filtro son los siguientes; para PM se utilizó una de ventana de tamaño 10 y para SGD un polinomio de segundo orden y una ventana de tamaño 7 en la etapa de suavizado.

Selección de Variables

Numerosos estudios han demostrado la utilidad de la selección de longitudes de onda características de cada elemento, para evitar información irrelevante o redundante mediante una amplia variedad de métodos (Algamal, 2019; Kawamura *et al.*, 2021; Zhang *et al.*, 2019). La selección de longitudes de onda mediante algoritmos genéticos (AG) optimiza las calibraciones multivariadas sin pérdida de capacidad de predicción (M. Yang *et al.*, 2019). En este estudio se utilizó un AG de variables fijas (Fei *et al.*, 2009), que fue acoplado con una regresión de mínimos cuadrados parciales (PLS) para la selección de un conjunto óptimo de variables espectrales características. Los valores de los parámetros del AG se establecieron de la siguiente manera: 1000 generaciones, tamaño de población de 50 cromosomas, tasa de mutación de 0.001 y 0.05 de cruzamiento.

La función objetivo a minimizar fue el error cuadrático medio de validación cruzada (RMSEcv).

Algoritmos de regresión

Regresión de Mínimos Cuadrados Parciales

La regresión de mínimos cuadrados parciales (PLS) es un modelo de enfoque lineal ampliamente utilizado en el análisis cuantitativo de espectros de reflectancia NIR en suelos (Ahmadi *et al.*, 2021). A través de una transformación lineal, PLS representa los datos espectrales originales a través de un conjunto de nuevas variables inferidas llamadas variables latentes. Estas son utilizadas para establecer el modelo de regresión lineal, además, son mutuamente ortogonales y no relacionadas, lo que elimina la multicolinealidad entre conjuntos de datos (Li *et al.*, 2019). En este estudio, el número de variables latentes se calculó mediante validación cruzada en el conjunto de entrenamiento del modelo.

Red Neuronal de Aprendizaje Extremo

Las redes de aprendizaje extremo (AE) es una red neuronal retroalimentada de una sola capa, los pesos de entrada y el sesgo son asignados de manera aleatoria, mientras que los pesos de salida se calculan con el método de mínimos cuadrados (Yang *et al.*, 2019). Su proceso se completa en una sola iteración, lo que le proporciona una alta velocidad de aprendizaje, además tiende a no tener problemas de sobreajuste o de tasas de aprendizaje inadecuados, para mayor detalle del algoritmo se pueden consultar trabajos anteriores (Vestergaard *et al.*, 2021; Yang *et al.*, 2019). El número de neuronas en la capa oculta debe ser establecido, sin embargo, en la práctica, no existe una teoría fija para determinar este parámetro. En el estudio se adoptó un método de prueba por pasos (H. Li *et al.*, 2019), estableciendo rangos del número de neuronas, luego se calculó el efecto del modelo bajo cada rango y finalmente selecciona el número óptimo en función del error mínimo. Con este método, el número de neuronas se estableció en 15.

Máquina de vectores de soporte

Los vectores de soporte (MVS) se utilizan ampliamente en la calibración de modelos espectrales para la predicción de nutrientes del suelo, como el nitrógeno y otros elementos (Guo et al., 2021; Hong et al., 2018; Wan et al., 2019; Y. Zhang et al., 2016).

MVS es un método de aprendizaje automático basado en funciones núcleo (Lijuan Liu et al., 2014) y perteneciente al enfoque de modelado no lineal. A través de estas funciones, el algoritmo MVS proyecta los datos de entrada originales en un espacio de características latentes en el modelado (V. Vapnik, 1998). Diversos estudios señalan la utilidad de la función de núcleo radial (RBF) (Thurnhofer-hemsi *et al.*, 2020) debido a su capacidad para manejar relaciones lineales y no lineales entre los espectros y los atributos del suelo, además de reducir la complejidad computacional (Ji-yong et al., 2013; Qi et al., 2018; S. Xu et al., 2018; M. Yang et al., 2019). El proceso de calibración de MVS implica la elección de la función de núcleo adecuada y la determinación de los parámetros óptimos del modelo, es decir, el parámetro de penalización (C) y el parámetro de escala de la función núcleo (γ). En este estudio se utilizó la función núcleo RBF, los parámetros C y γ se establecieron en 1000 y 0.001 respectivamente mediante validación cruzada del modelo.

Bosques Aleatorios

Este enfoque de aprendizaje de automático corresponde al aprendizaje de ensamble, está basado en dos conceptos clave; los árboles de decisión (Hastie *et al.*, 2005) y el concepto de empaquetado o Bagging (Brieman, 1996). Básicamente, bosques aleatorios (BA) construye un conjunto de árboles no correlacionados y luego promedia sus salidas para obtener modelos robustos. La idea en BA es reducir la varianza del conjunto al reducir la correlación entre los árboles. Esto se logra a través de árboles de decisión contruidos a partir de la selección aleatoria de subconjuntos de variables de entrada. BA ha sido ampliamente utilizado en análisis espectral para ciencias del suelo, su desempeño ha sido adecuado en diversos estudios relacionados con la disposición de nutrientes (Ding *et al.*, 2018; R. Vestergaard *et al.*, 2021). Con

base en el RMSEcv, el número de árboles que produjo mejores resultados para esta investigación fue de 300.

Incremento de datos

Redes neuronales adversarias

Para mejorar la calidad y cantidad de los datos para entrenamiento de los algoritmos y analizar el efecto de este aumento, se propuso el uso de redes neuronales generativas GAN, como una solución para la escasez de cantidad y diversidad de los datos espectrales. GAN se ha implementado para agregar datos de entrenamiento en varios estudios relacionados con la percepción remota y espectroscopía (Lv *et al.*, 2021; Wambugu *et al.*, 2021; Wu *et al.*, 2021). Así mismo, se han empleado para la mejora de la predicción de nutrientes de suelos en trabajos como el de (Jiang *et al.*, 2023).

GAN (Goodfellow *et al.*, 2020), es un tipo de aprendizaje profundo que consiste en dos redes neuronales (generador y discriminador) que se entrenan en paralelo. El generador (G) se entrena para producir datos sintéticos o artificiales, mientras que el discriminador (D) se entrena para distinguir los datos artificiales de los reales. El entrenamiento consta de un número determinado de iteraciones o épocas, en cada época, tanto G como D trabajan con la función objetivo de entropía cruzada (4.1) pero de forma opuesta. Esta función mide la discrepancia entre la distribución de probabilidad de las muestras generadas y la distribución de probabilidad de las muestras reales.

$$\arg \min_G \max_D \mathbf{E}_{z,x} [\log D(G(z)) + \log(1 - D(x))] \quad (4.1)$$

Donde $G(z)$ son los datos generados, $D(G(z))$, $D(x)$ y $(1 - D(x))$ representan las probabilidades estimadas de verdaderos negativos, falsos negativos y verdaderos positivos del clasificador D respectivamente.

La finalidad de G es minimizar la entropía cruzada. Inicialmente, utiliza ruido aleatorio con distribución Gaussiana o normal (z), que se va transformando en cada época, generando muestras cada vez más semejantes a las reales y complicando la tarea de clasificación de D . Por otro lado, D busca maximizar esta

función objetivo, mejorando la probabilidad de identificar verdaderos positivos (reales) y verdaderos negativos (artificiales). Esta competencia adversarial entre G y D permite mejorar gradualmente la calidad de las muestras generadas. Debido al costo computacional que implica este proceso en espectros y al equipo computacional disponible, en este estudio el valor máximo de épocas se estableció en 300.

Efecto del aumento de datos en los modelos.

Para investigar el efecto del aumento de datos artificiales en los modelos espectrales de NA, se comparó la precisión de predicción de cada uno, realizando aumentos graduales (niveles) en el conjunto inicial de espectros reales mediante la adición de espectros generados con GAN. En cada nivel, se añadieron 61 espectros artificiales para entrenamiento y los modelos se utilizaron para predecir un mismo conjunto de validación. Los niveles de aumento utilizaron 95 muestras reales más n muestras artificiales originando siete conjuntos adicionales, para cada conjunto C_i , la cantidad de muestras añadidas n_i fue la siguiente: $n_1 = 0, n_2 = 61, n_3 = 122, n_4 = 183, n_5 = 244, n_6 = 305, n_7 = 366, n_8 = 427$.

4.2.4. Evaluación

Métricas de predicción

La capacidad de predicción de los modelos se evaluó calculando el coeficiente de determinación de la predicción (R^2), la raíz del error cuadrático medio de predicción (RMSEp) y la desviación predictiva residual (RPD). Valores predictivos altos para R^2 y bajos para RMSEp indican modelos más apropiados. RPD es el cociente entre la desviación estándar de los datos de validación y el RMSEp (Bellon-Maurel *et al.*, 2010). Para aplicaciones agrícolas: $2 \leq RPD \leq 3$ indican una buena capacidad predictiva, $1.5 \leq RPD \leq 2$ son modelos perfectibles y $RPD < 1.5$ es un indicativo de mala capacidad predictiva (D'Acqui *et al.*, 2010). Cada modelo de regresión fue corrido 10 veces, se reportan los valores promedios de las métricas.

2.4.2. Análisis de los datos artificiales de nitrógeno

Para los datos artificiales de nitrógeno se utilizaron diagramas de caja para reflejar la similitud de los datos de nitrógeno generados por GAN con los datos reales. Se compararon los cuartiles: máximo, mínimo, medio, mediano y superior e inferior de los datos. Finalmente, los datos se contrastaron utilizando las diferencias numéricas entre el máximo, el mínimo, la media, la mediana y la desviación estándar de ambos conjuntos.

2.4.3. Análisis de los espectros artificiales.

Para el estudio y comparación de los datos espectrales artificiales, se utilizó el análisis por componentes principales (PCA). El principio básico de PCA es convertir los datos originales de dimensión m en un nuevo conjunto de variables ortogonales k -dimensionales llamadas componentes principales (Wold et al., 1987). La reducción de la dimensionalidad mediante PCA permite evaluar la similitud y la diversidad de los espectros generados en comparación con los originales (Jiang et al., 2023). En este estudio, los datos espectrales tienen una dimensiones de $m = 4615$, los cuales se redujeron a $k = 2$ componentes para simplificar la comparación entre conjuntos. Luego, la distribución de datos generados y reales se compararon gráficamente observando la forma de las curvas y sus características (media y desviación estándar) y diagramas de violín divididos.

4.3. RESULTADOS Y DISCUSIÓN

4.3.1 Descripción estadística de los datos de nitrógeno reales y artificiales

La estadística descriptiva de los datos de nitrógeno aprovechable se muestra en el Cuadro 4.1, hay similitudes en las medidas de tendencia central entre los datos reales y los datos artificiales, ya que las medias y medianas tienen variaciones porcentuales menores al 5%. La forma de las distribuciones es similar, pues ambas distribuciones tienen asimetría positiva. Sin embargo, se pueden apreciar diferencias más grandes en otras medidas de dispersión. La desviación estándar de los datos reales es 46% mayor que la de los datos artificiales, además el coeficiente de variación difiere en una proporción similar, indicando que los valores artificiales tienden a estar más cerca de su media en comparación con

los reales. Esto indica que los datos reales presentan una mayor dispersión en la concentración de nitrógeno, lo cual también se refleja en un rango intercuartílico más amplio y en la presencia de valores atípicos por encima del adyacente superior (Figura 4.3).

En los datos reales, se tiene una asimetría positiva de 0.98, lo que se refleja en una extensión mayor del adyacente superior. Esto también se observa en la distribución de los datos artificiales, pero en menor proporción.

Cuadro 4.2. Estadística descriptiva de los datos de nitrógeno aprovechable (NA) real y artificial

Estadístico	NA real	NA art.
	(mg kg ⁻¹)	
Media	25.60	25.71
Desv.est	7.64	5.21
Asimetría	0.98	0.68
Min	13.34	17.96
Max	47.20	37.50
CV*	29.85 %	20.30 %
Mediana	22.94	23.90

*CV=coeficiente de variación

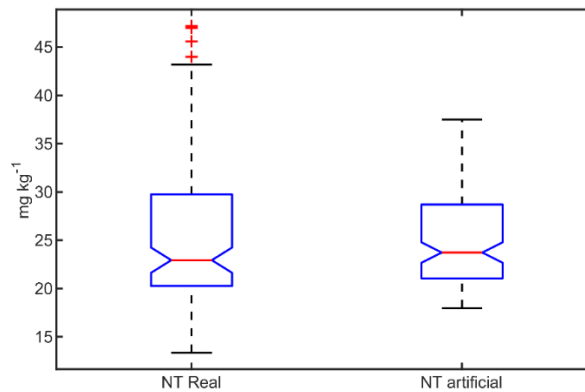


Figura 4.3. Diagrama de caja de las distribuciones del nitrógeno aprovechable (real) y el generado artificialmente.

En general, los datos artificiales muestran una distribución similar a los datos reales en términos de tendencia central coincidiendo en este sentido con los resultados de Jiang *et al.*, (2023), pero con una menor variabilidad. Esto podría ser un efecto de simplificación o suavizado de la distribución original durante el proceso de generación en GAN producto de falta de diversidad y sesgo de los datos de entrenamiento (Aggarwal *et al.*, 2021).

4.3.2. Análisis de los espectros reales y artificiales

Espectros reales

Se obtuvieron espectros de reflectancia en el rango espectral NIR (800-2700 nm) con 4615 variables. Los espectros sin procesar y preprocesados se presentan en la Figura 4.4. La tendencia general de los espectros NIR fue similar, lo cual se puede atribuir a la presencia de las mismas propiedades espectralmente activas en todas las muestras recolectadas en el campo (S. Xu et al., 2018). Tanto en los espectros originales (Figura 4.4a) como en los derivados (Figura 4.4b), se observó la presencia de ruido de alta y baja frecuencia, principalmente en el rango espectral 800-1000 nm, lo cual coincide a su vez, con un grado de pendiente notable en la misma región de la línea base del espectro, esto es causado comúnmente por factores no sistemáticos en el equipo de medición espectral (Ben Dor *et al.*, 2015).

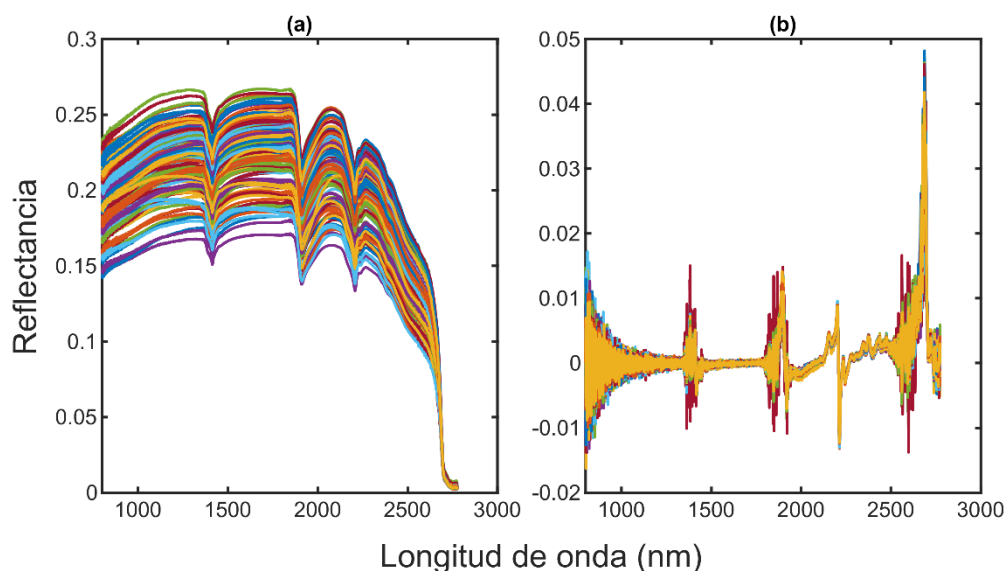


Figura 4.4. Espectros originales(a), Primera derivada de la reflectancia (b)

Las muestras recolectadas en campo exhibieron tres características de absorción prominentes en LO de aproximadamente 1400, 1900 y 2200 nm. Las dos primeras regiones de absorción se atribuyen a la presencia de moléculas de agua atrapadas en la red cristalina del suelo. En 1400 nm la absorción es causada por los armónicos de estiramiento en enlaces de oxígeno e hidrógeno (O-H), en 1900

nm estos armónicos se combinan con armónicos de flexión H-O-H (Clark *et al.*, 1990). Estas características reflejaron las condiciones de humedad de campo en las muestras y podrían haber causado diferencias espectrales debido a las interacciones entre el agua y otros componentes (Reeves, 2010). El pico de absorción en aproximadamente 2200 nm está relacionado con la absorción de Al-OH por arcillas en el suelo (Clark & Roush, 1984), lo que sugiere una alta proporción de partículas de tamaño pequeño (limo y arcilla) en las muestras.

Análisis PCA de los espectros

Los espectros obtenidos en laboratorio se introdujeron en las redes GAN para su entrenamiento, como resultado se obtuvo una matriz de 427 x 4616, donde la primera dimensión es el número de muestras artificiales y la segunda el número de LO de los espectros, más una columna con los datos de nitrógeno generados. Los espectros reales, así como los artificiales, se redujeron a dos dimensiones usando PCA, es decir, se seleccionaron los primeros dos componentes principales (CP1 y CP2). El Cuadro 4.2 muestra la tasa de contribución de CP1 y CP2 correspondiente a los espectros reales y los generados con GAN. Los resultados sugieren que los dos conjuntos de datos tienen una estructura similar en términos de la importancia de los dos primeros CP, con un aporte total del 98% para ambos conjuntos. Sin embargo, existe una diferencia del 4% en la contribución individual de los componentes, lo que puede indicar ciertas diferencias en la estructura subyacente de los datos.

Cuadro 4.2. Resultados del análisis de componentes principales

Conjunto espectral	Contribución de los CP (%)	
	CP1	CP2
Real	94.67	4.47
GAN	90.43	8.29

Para visualizar estas similitudes y discrepancias, la distribución de densidad de probabilidad de los datos bidimensionales se trazó mediante diagramas de violín divididos, los resultados se muestran en la Figura 4.5a. La similitud entre las formas de ambas distribuciones sugiere que ambos conjuntos tienen características o patrones subyacentes comunes. Sin embargo, se observan

ciertas discrepancias en las longitudes de las distribuciones, debido a diferencias de variabilidad, siendo mayor la de los datos reales. Además, las variaciones en la mediana y la distancia intercuartil entre conjuntos, también indican una diferencia sistemática en el valor central de los datos y valores extremos de cada conjunto. Asimismo, lo anterior se refleja en la Figura 4.5b, donde se presenta una comparación grafica entre los espectros promedio de ambos conjuntos, así como de sus desviaciones estándar.

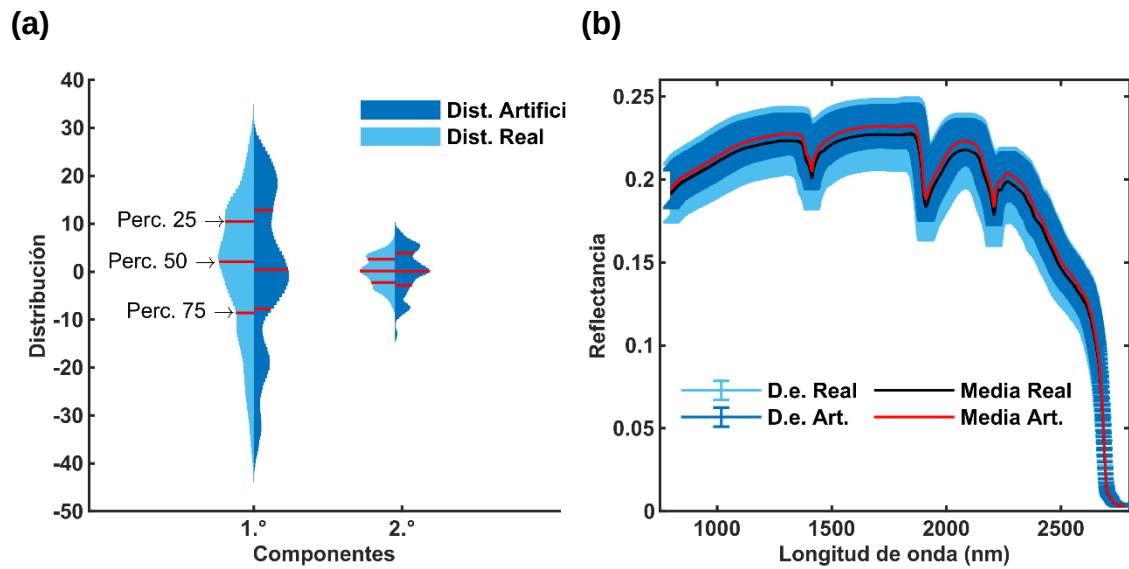


Figura 4.5. Diagramas comparativos de los dos primeros componentes principales de la distribución espectral (a) y comparación grafica de la reflectancia media y su desviación estándar (b) para los conjuntos de datos reales y artificiales.

La forma general de los espectros artificiales coincide con los espectros originales. Se observan los mismos puntos críticos de absorción en 1499, 1900 y 2200 nm en ambos conjuntos, pero con una dispersión menor a la real y variaciones en la reflectancia promedio.

4.3.3. Impacto del aumento de datos en los algoritmos de regresión

Para realizar las regresiones, se seleccionaron 10 longitudes de onda mediante GA-PLS, estas variables se localizaron principalmente en el rango espectral de 900-1010 nm, una zona espectral con fuertes relaciones con el contenido de nitrógeno en suelos (Mukherjee y Laskar, 2019; Yao Zhang *et al.*, 2019). Así

mismo, longitudes de onda entre 1315 y 1400 nm también fueron relacionadas con el nitrógeno en esta investigación, lo que se refuerza con estudios previos como el de Zhou *et al.* (2018).

En la Figura 4.6 se muestran los cambios en el RMSEp y el coeficiente R^2 de cada modelo, las métricas de evaluación completas se proporcionan en el Cuadro 4.3. Para cada grafico en la Figura 4.6, el primer punto representa el desempeño del modelo entrenado con el conjunto original y cada punto posterior corresponde a las métricas obtenidas con siete niveles de aumento de datos en el conjunto de entrenamiento. Como se puede observar, cuando sólo se usaron datos reales para la calibración de los modelos, el R^2 de la mayoría de los modelos fue baja, siendo el mejor para MVS con un R^2 de 0.63 y RMSEp de 2.59. Sin embargo, el rendimiento de los modelos después del aumento de datos reveló que los espectros artificiales contribuyeron de manera positiva en el aprendizaje de los algoritmos, reforzando lo reportado por Jiang *et al.* (2023). A medida que aumentaba el número de muestras, el R^2 mejoró y el RMSEp disminuyó, sin embargo, en ciertos niveles de aumento, el rendimiento de los modelos no varió o se deterioró. Para determinar el mejor nivel de aumento de datos en base a los coeficientes de determinación R^2 y el error de predicción RMSEp, es necesario encontrar un equilibrio entre la mejora del rendimiento y la complejidad añadida por los datos artificiales.

En esta investigación, hubo niveles de aumento en los que los algoritmos tuvieron comportamientos clave. Por ejemplo, el cuarto nivel fue importante en el aprendizaje de PLS y AE, los aumentos de más de 244 muestras no contribuyeron considerablemente en la reducción del RMSEp de PLS, y su R^2 se mantuvo en valores muy similares al modelo sin aumento (Figura 4.6a).

Por otro lado, a pesar de la tendencia de reducción del RMSEp en niveles posteriores, después del nivel 4, AE no mejoró su capacidad explicativa de la variabilidad total, pues se observó una estabilización del R^2 (Figura 4.6b).

Tanto para MVS como para BA (Figuras 4.6c y 4.6d), se observó un comportamiento más consistente durante el aumento de datos de entrenamiento con relación a los modelos anteriores. Los valores de R^2 y RMSEp para ambos

modelos indicaron una mejora continua en su aprendizaje y capacidad de generalización. Además, se notó una relación directa entre el rendimiento del modelo y el coeficiente de determinación R^2 , ya que ambos aumentaron y disminuyeron en sincronía.

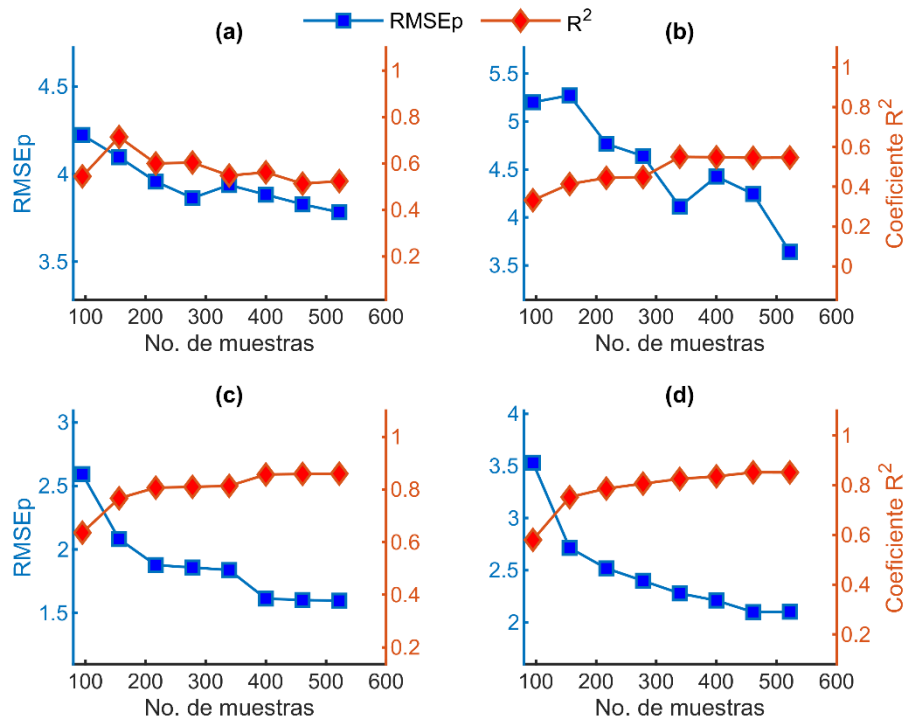


Figura 4.6. Efecto del aumento de datos en los parámetros de predicción de los modelos de regresión PLS(a), AE (b), SVM(c) y BA (d).

El hecho de que el RMSE y el R^2 no cambiaran o se deterioraran a partir de niveles de aumento específicos, 5 y 6 para los casos de MVS y BA respectivamente, indica que los modelos alcanzaron un punto de estabilidad y no se beneficiaron significativamente de la adición de más datos. En este punto, la adición de datos pudo introducir cierto nivel de ruido o ambigüedad, lo que se refleja en una estabilización o deterioro de las métricas. Este comportamiento también fue observado en los resultados de Jiang *et al.*, (2023) con redes neuronales convolucionales con espectros de suelos.

Cuadro 4.3. Comparación del desempeño de los modelos de calibración a través del aumento gradual de daos de entrenamiento

Métrica	Algoritmo	Real (95)	Niveles de aumento de datos						
			1(+61)	2(+122)	3(+183)	4(+244)	5(+305)	6(+366)	7(+427)

RMSEt	PLS	3.76	2.78	3.192	3.104	3.476	3.402	3.54	3.51
	MVS	1.220	0.941	0.840	0.786	0.763	0.763	0.740	0.721
	AE	3.851	2.862	3.628	4.097	4.131	4.317	4.233	4.269
	BA	3.038	2.810	2.462	2.100	2.144	2.072	2.150	2.102
RMSEp	PLS	4.224	4.097	3.958	3.864	3.9381	3.883	3.828	3.783
	MVS	2.593	2.082	1.877	1.858	1.838	1.613	1.602	1.597
	AE	5.201	5.273	4.768	4.641	4.114	4.427	4.545	3.642
	BA	3.531	2.714	2.518	2.399	2.279	2.210	2.099	2.104
R²	PLS	0.545	0.714	0.6	0.604	0.548	0.562	0.513	0.524
	MVS	0.636	0.767	0.807	0.810	0.814	0.857	0.859	0.860
	AE	0.332	0.414	0.445	0.448	0.550	0.548	0.545	0.547
	BA	0.580	0.752	0.786	0.806	0.825	0.835	0.852	0.851
RPD	PLS	1.052	1.362	1.401	1.441	1.416	1.434	1.453	1.47
	MVS	1.782	2.214	2.515	2.550	2.575	2.890	2.900	2.906
	AE	1.084	1.057	1.188	1.205	1.396	1.282	1.322	1.546
	BA	1.571	2.043	2.202	2.312	2.434	2.510	2.642	2.638

Además, teniendo en cuenta el efecto de los datos agregados en el ajuste y el error y considerando el puntaje RPD máximo alcanzado (Cuadro 4.3), se puede establecer un nivel adecuado de aumento de datos para cada modelo. En este estudio se determinó que para PLS, la cantidad óptima de datos fue de 122, sin embargo el nivel de precisión (RPD) fue bajo comparado con trabajos como los de Y. Li *et al.*, (2019) y Morellos *et al.*, (2016), cuyos valores RPD fueron superiores a 2 con un numero de muestras similar. Esto se puede atribuir a una menor calidad y representatividad de los datos y un nivel de ruido superior en este estudio, lo cual también se observó para el caso de EA, donde con 522 muestras como la mejor cantidad de datos no se alcanzaron puntuaciones RPD altas como en el estudio de H. Li *et al.*(2019) quien utilizó menos datos. En el caso de MVS, el rendimiento máximo se alcanzó con 400 muestras, alcanzando niveles de precisión (RPD) que coinciden con el estudio de Xu *et al.* (2018) con un numero de muestras similar en predicción de nitrógeno. Para BA el sexto nivel con 461 muestras alcanzó la mejor estabilidad y equilibrio entre RMSEp, R² y RPD.

4.3.4 Comparación del desempeño de los modelos de calibración

Precisión de calibración y de predicción

En general, una diferencia pequeña entre el error de calibración RMSE_t y el error de predicción RMSE_p sugiere que el modelo ha logrado un equilibrio entre su capacidad de ajustarse a los datos de entrenamiento y su capacidad de generalizar con nuevos datos. Esto es deseable, ya que indica una mayor confiabilidad y robustez en las predicciones (Montesinos-López *et al.*, 2022).

Dentro de los límites de capacidad de aprendizaje de cada modelo, el modelo PLS muestra una consistencia en su adaptación a nuevos datos a medida que aumenta la cantidad de datos artificiales. Se observaron diferencias pequeñas entre los errores RMSE y RMSE_p en la mayoría de los niveles de aumento.

Para MVS, la diferencia inicial mayor entre los errores de los primeros cuatro niveles de aumento, indica que el modelo podría haber tenido un grado de sobreajuste mayor en comparación con PLS. Sin embargo, la reducción notable en esta diferencia en niveles posteriores sugiere que el modelo logró mejorar su capacidad de generalización a medida que se agregaban más datos.

Con AE las diferencias entre los errores en los primeros dos niveles de aumento también fueron superiores a las observadas con PLS. Sin embargo, a partir de esos niveles, las diferencias se redujeron notablemente. Al igual que PLS, se observó una mayor diferencia de errores en el primer nivel de aumento. Esto podría indicar una mayor sensibilidad inicial de los modelos a datos que difieren en su estructura subyacente del conjunto original.

Finalmente, con BA la diferencia entre errores fue mínima en todos los niveles de aumento, lo que indica una capacidad de generalización superior. Esto sugiere que el modelo BA pudo aprovechar eficazmente la información de los datos artificiales para mejorar su capacidad de predicción.

Ajustes de predicción

Los cuatro métodos multivariantes proporcionaron una precisión de predicción máxima en diferentes niveles de aumento de datos. Los diagramas de dispersión del contenido de NA pronosticado frente a las medidas obtenidas en laboratorio

se presentan en la Figura 4.7. El mejor resultado se obtuvo con el modelo MVS, en el nivel 5 con 400 muestras, correspondiente a 95 reales y 305 artificiales (Figura 4.7a). Con un R^2 de 0.857 y un RPD de 2.89, el modelo parece tener una buena capacidad de ajuste en nuevos datos. SVM obtuvo los menores índices de error y una buena precisión en las predicciones, confirmando su alta viabilidad previamente reportada por Zhang *et al.* (2016) para nitrógeno en suelos, sobre los métodos lineales.

El modelo BA (Figura 4.7b), también mostró un alto rendimiento, con un R^2 de 0.859 y un RPD de 2.64 con 461 muestras (nivel 6). Además, después de MVS, BA fue el modelo con menores índices de error, mostrando una buena capacidad de ajuste y precisión en las predicciones. Por otro lado, tanto AE como PLS en este estudio mostraron un menor ajuste y una capacidad predictiva limitada (Figuras 4.7c y 4.7d respectivamente). Con los errores más grandes, R^2 de 0.6 y 0.54 y un RPD de 1.4 y 1.54 respectivamente son los métodos con el desempeño más bajo y con menor viabilidad de aplicación práctica.

Las diferencias en los resultados obtenidos con los métodos evaluados se deben a la diversidad en las estructuras básicas de cada algoritmo. MVS y BA, al ser métodos no lineales más profundos, tienen la capacidad de capturar relaciones no lineales más complejas en los datos, lo que pudo permitirles un mejor desempeño de predicción en comparación con AE y PLS.

De acuerdo con Liu *et al.*, 2014, MVS utiliza una estrategia de mapeo no lineal de los datos en un hiperplano de alta dimensión mediante el uso de funciones de núcleo, lo que le permite capturar relaciones no lineales de manera más efectiva. Por otro lado, BA es un conjunto de árboles de decisión, donde cada árbol es entrenado con una muestra aleatoria de los datos, lo que crea diversidad en el conjunto de árboles. Esta diversidad y combinación de árboles individuales en BA permiten capturar relaciones no lineales y las interdependencias en las variables predictoras. Además, las predicciones finales se obtienen mediante una combinación de las predicciones de cada árbol, lo cual permite reducir el sesgo y la varianza del modelo. Esto resulta en una mejor capacidad de generalización como menciona Cootes *et al.* (2012).

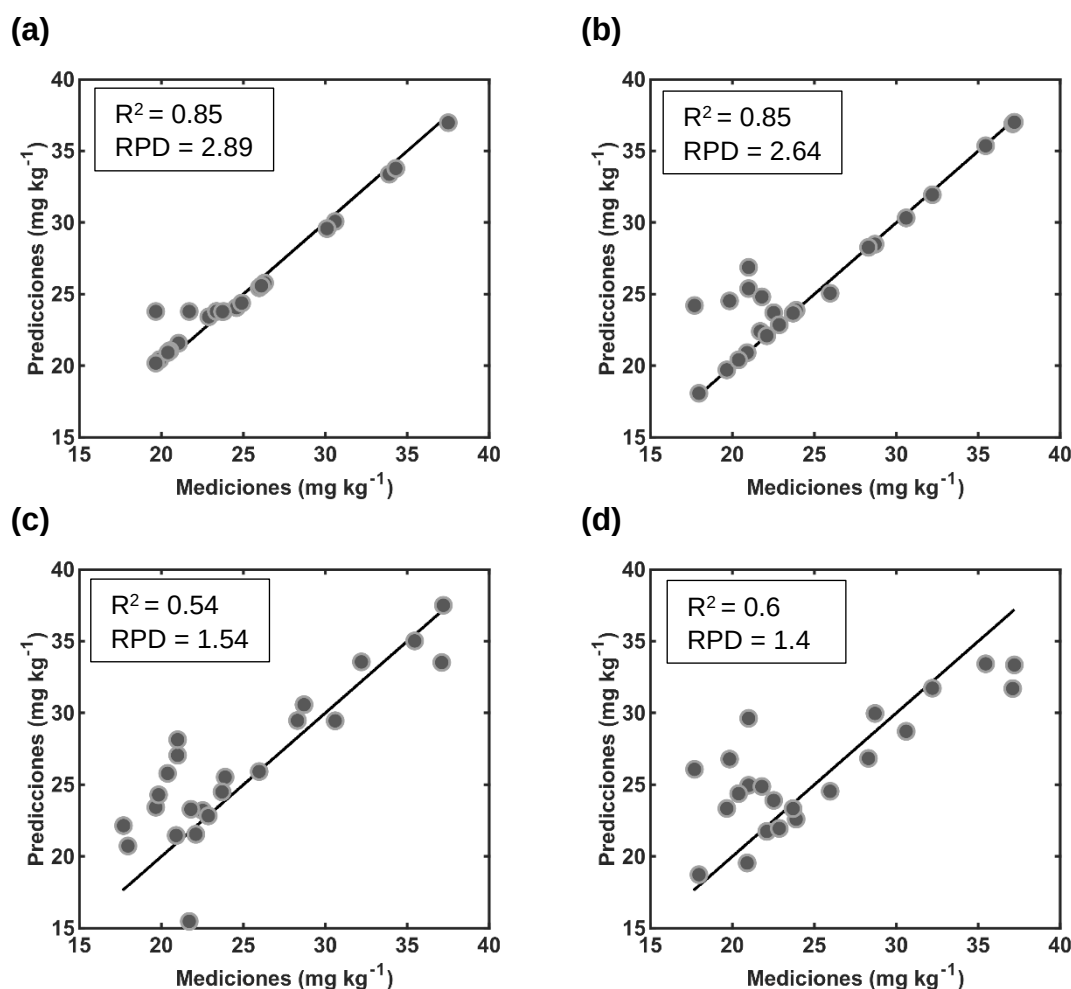


Figura 4.7. Mejores ajustes de los modelos con aumento de datos; (a) MVS en el quinto nivel, (b) BA en el sexto, (c) AE en el séptimo, (d) PLS en el segundo.

Por otro lado, aunque AE también es un enfoque no lineal, tiene una estructura más simple en comparación con MVS y BA. AE utiliza una única capa oculta de neuronas y se basa en una estrategia de mapeo no lineal de las entradas mediante la función de activación. Sin embargo, la selección aleatoria de los pesos de conexión y la falta de ajuste iterativo (Huang *et al.*, 2006), pudieron haber limitado su capacidad para capturar relaciones no lineales complejas.

Asimismo, a diferencia de los otros enfoques, PLS es un método lineal multivariante que busca maximizar la covarianza entre las variables predictoras y la variable objetivo, asumiendo una relación lineal entre ellas (Helland, 1990).

Esto puede limitar su capacidad para modelar las relaciones no lineales complejas de los datos.

Estos resultados coinciden con los reportados en estudios espectrales para predicción de NA en suelos como los de Morellos *et al.* (2016) y Xu *et al.* (2018). Donde los métodos multivariante con enfoque no lineal como MVS superan el desempeño de modelos de naturaleza lineal como PLS. AE a pesar de presentar el peor rendimiento de los métodos no lineales también fue mejor que el método lineal lo cual se respalda con los hallazgos de Li *et al.* (2019). Además, el buen rendimiento de RF en la predicción de nitrógeno de suelos, también es consistente con los estudios de Ding *et al.* (2018) y R.-J. Vestergaard *et al.* (2021), mostrando su viabilidad para este tipo de aplicaciones.

4.4. CONCLUSIONES

Los resultados sugieren que las relaciones entre la reflectancia del suelo y el contenido de nitrógeno aprovechable pueden ser muy complejas y no lineales. Los modelos de vectores de soporte y bosques aleatorios demostraron el mejor rendimiento predictivo en términos de las métricas de evaluación empleadas. Estos modelos fueron capaces de capturar relaciones no lineales y lograron un buen ajuste a los datos. Aunque la red de aprendizaje extremo es un enfoque no lineal, su estructura más simple pudo haber limitado su capacidad para capturar relaciones no lineales complejas. Los modelos lineales de mínimos cuadrados parciales no alcanzaron el mismo nivel de precisión.

En general, se reafirma la utilidad de modelos generativos como GAN, como una alternativa viable para la ampliación y mejora de bases de datos espectrales de reflectancia NIR. El aumento de datos artificiales en el conjunto de entrenamiento resultó en mejoras significativas en el rendimiento de los modelos, pero el impacto específico varió entre ellos.

En investigaciones futuras, sería valioso investigar y desarrollar modelos no lineales más avanzados y sofisticados. Esto podría incluir la exploración de técnicas de aprendizaje profundo, redes neuronales más complejas y métodos más avanzados de inteligencia artificial que puedan capturar las relaciones no lineales. Además, sería interesante explorar y comparar otros tipos de modelos

generativos. Esto podría involucrar la investigación y comparación con modelos como los codificadores automáticos variacionales (VAEs) u otros enfoques generativos para mejorar el rendimiento de los modelos.

Referencias del capítulo

- Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1). <https://doi.org/10.1016/j.jjime.2020.100004>
- Ahmadi, A., Emami, M., Daccache, A., & He, L. (2021). Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis. *Agronomy*, 11(3). <https://doi.org/10.3390/agronomy11030433>
- Albawi, S., & Mohammed, T. A. (2017). Understanding of a Convolutional Neural Network. *IEEE*. <https://ieeexplore.ieee.org/abstract/document/8308186>
- Algarni, Z. Y. (2019). Variable selection in count data regression model based on firefly algorithm. *Statistics, Optimization and Information Computing*, 7(2), 520–529. <https://doi.org/10.19139/soic.v7i2.566>
- Alibabaei, K., Gaspar, P. D., Lima, T. M., Campos, R. M., Girão, I., Monteiro, J., & Lopes, C. M. (2022). A Review of the Challenges of Using Deep Learning Algorithms to Support Decision-Making in Agricultural Activities. *Remote Sensing*, 14(3), 1–43. <https://doi.org/10.3390/rs14030638>
- Andrade, R., Faria, W. M., Silva, S. H. G., Chakraborty, S., Weindorf, D. C., Mesquita, L. F., Guilherme, L. R. G., & Curi, N. (2020). Prediction of soil fertility via portable X-ray fluorescence (pXRF) spectrometry and soil texture in the Brazilian Coastal Plains. *Geoderma*, 357(July 2019), 113960. <https://doi.org/10.1016/j.geoderma.2019.113960>
- Araújo, S. R., Wetterlind, J., Demattê, J. A. M., & Stenberg, B. (2014). Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. *European Journal of Soil Science*, 65(5), 718–729. <https://doi.org/10.1111/ejss.12165>
- Arcos, J. M., Cruz Ortiz, M., Villahoz, B., & Sarabia, L. A. (1997). Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: Application on copper and zinc mixture. *Talanta*, 59(2), 311–317. [https://doi.org/10.1016/S0039-9140\(02\)00505-2](https://doi.org/10.1016/S0039-9140(02)00505-2)
- Armenise, E., Redmile-Gordon, M. A., Stellacci, A. M., Ciccarese, A., & Rubino, P. (2013). Developing a soil quality index to compare soil fitness for agricultural use under different managements in the mediterranean environment. *Soil and Tillage Research*, 130, 91–98. <https://doi.org/10.1016/j.still.2013.02.013>
- Atasoy, S., Mateus, D., Georgiou, A., Navab, N., & Yang, G. Z. (2011). Wave interference for pattern description. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6493 LNCS(PART 2), 41–54. https://doi.org/10.1007/978-3-642-19309-5_4
- Baek, J. Y., Yoo, Y. S., & Bae, S. H. (2019). Adversarial Learning With Knowledge of Image Classification for Improving GANs. *IEEE Access*, 7, 56591–56605. <https://doi.org/10.1109/ACCESS.2019.2913697>
- Barber, D. (2020). *Bayesian Reasoning and Machine Learning*.
- Barefoot, A., Murphy, J., & Aizawa, H. (2003). *Handbook of Residue Analytical Methods for Agrochemicals* (P. W. Lee (ed.); Wiley).

- 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>
- Bartholdi, E., & Ernst, R. R. (1973). Fourier spectroscopy and the causality principle. *Journal of Magnetic Resonance* (1969), 11(1), 9–19. [https://doi.org/10.1016/0022-2364\(73\)90076-0](https://doi.org/10.1016/0022-2364(73)90076-0)
- Bellon-Maurel, V., Fernandez-ahumada, E., Palagos, B., Roger, J., & Mcbratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry*, 29(9), 1073–1081. <https://doi.org/10.1016/j.trac.2010.05.006>
- Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., & Sommer, S. (2009). Using Imaging Spectroscopy to study soil properties. *Remote Sensing of Environment*, 113(SUPPL. 1), S38–S55. <https://doi.org/10.1016/j.rse.2008.09.019>
- Ben-dor, E., & Epema, G. F. (2014). *Soil Reflectance. January 1999*.
- Ben Dor, E., Ong, C., & Lau, I. C. (2015). Reflectance measurements of soils in the laboratory: Standards and protocols. *Geoderma*, 245–246, 112–124. <https://doi.org/10.1016/j.geoderma.2015.01.002>
- Brame, C. J. (2007). A guide To active learning -Vanderbilt white paper. *Vanderbilt University Center for Teaching*. <https://cft.vanderbilt.edu/active-learning/>
- Brieman, L. (1996). Bagging predictors. *Machine Learning*, 24(3), 123–140. <https://doi.org/10.3390/risks8030083>
- Bronick, C. J., & Lal, R. (2021). Soil structure and management : a review. *Geoderma*, 124(2005), 3–22. <https://doi.org/10.1016/j.geoderma.2004.03.005>
- Brunton, S. L., & Kutz, J. N. (2019). *Data-Driven Science and Engineering* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108380690>
- Buchröller, W., & Tye, S. H. H. (1980). Vibrational states in the spectroscopy. *Physical Review Letters*, 44(13), 850–853. <https://doi.org/10.1103/PhysRevLett.44.850>
- Buddenbaum, H., & Steffens, M. (2012). *The Effects of Spectral Pretreatments on Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy. 2012*. <https://doi.org/10.1155/2012/274903>
- Burger, J., & Geladi, P. (2007). Spectral pre-treatments of hyperspectral near infrared images: Analysis of diffuse reflectance scattering. *Journal of Near Infrared Spectroscopy*, 15(1), 29–37. <https://doi.org/10.1255/jnirs.717>
- Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J., & Smaling, E. M. A. (2012). Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. *Geoderma*, 183–184, 41–48. <https://doi.org/10.1016/j.geoderma.2012.03.011>
- Cameron, E. K., Martins, I. S., Lavelle, P., Mathieu, J., Tedersoo, L., Gottschall, F., Guerra, C. A., Hines, J., Patoine, G., Siebert, J., Winter, M., Cesarz, S., Delgado-Baquerizo, M., Ferlian, O., Fierer, N., Kreft, H., Lovejoy, T. E., Montanarella, L., Orgiazzi, A., ... Eisenhauer, N. (2018). Global gaps in soil biodiversity data. *Nature Ecology and Evolution*, 2(7), 1042–1043. <https://doi.org/10.1038/s41559-018-0573-8>
- Cao, Z., Kühn, P., He, J.-S., Bauhus, J., Guan, Z.-H., & Scholten, T. (2022). Calibration of Near-Infrared Spectra for Phosphorus Fractions in Grassland Soils on the Tibetan Plateau. *Agronomy*, 12(4), 783. <https://doi.org/10.3390/agronomy12040783>
- Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1), 103–112. <https://doi.org/10.1016/j.chemolab.2004.12.011>

- Ciani, A., Goss, K. U., & Schwarzenbach, R. P. (2005). Light penetration in soil and particulate minerals. *European Journal of Soil Science*, 56(5), 561–574.
<https://doi.org/10.1111/j.1365-2389.2005.00688.x>
- Clark, R. N., King, T. V. V., Klejwa, M., Swayze, G. A., & Vergo, N. (1990). High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research*, 95(B8).
<https://doi.org/10.1029/jb095ib08p12653>
- Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, 89(B7), 6329–6340.
<https://doi.org/10.1029/JB089iB07p06329>
- Cootes, T. F., Ionita, M. C., Lindner, C., & Sauer, P. (2012). Robust and accurate shape model fitting using random forest regression voting. In A. Fitzgibbon (Ed.), *Lecture Notes in Computer Science: Vol. 7578 LNCS* (Issue PART 7, pp. 278–291). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33786-4_21
- Coutinho, M. A. N., Alari, F. de O., Ferreira, M. M. C., & Amaral, L. R. d. (2019). Influence of soil sample preparation on the quantification of NPK content via spectroscopy. *Geoderma*, 338(July 2018), 401–409. <https://doi.org/10.1016/j.geoderma.2018.12.021>
- D. C. Slaughter, M. G. Pelletier, & S. K. Upadhyaya. (2001). Sensing Soil Moisture Using Nir Spectroscopy. *Applied Engineering in Agriculture*, 17(2), 241–247.
<https://doi.org/10.13031/2013.5449>
- D’Acqui, L. P., Pucci, A., & Janik, L. J. (2010). Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. *European Journal of Soil Science*, 61(6), 865–876.
<https://doi.org/10.1111/j.1365-2389.2010.01301.x>
- Dhaliwal, S. S., Naresh, R. K., Mandal, A., Walia, M. K., Gupta, R. K., Singh, R., & Dhaliwal, M. K. (2019). Effect of manures and fertilizers on soil physical properties, build-up of macro and micronutrients and uptake in soil under different cropping systems: a review. *Journal of Plant Nutrition*, 42(20), 2873–2900. <https://doi.org/10.1080/01904167.2019.1659337>
- Ding, J., Yang, A., Wang, J., Sagan, V., & Yu, D. (2018). Machine-learning-based quantitative estimation of soil organic carbon content by VIS/NIR spectroscopy. *PeerJ*, 2018(10), 1–24.
<https://doi.org/10.7717/peerj.5714>
- Domingos, P. (2015). *The Master Algorithm*. Perseus Books Group.
- Donald, P., Gary, L., Kriz, G., & Vyvyan, J. (2017). Introduction to Spectroscopy. In *Spectroscopic Methods in Food Analysis* (5th ed.). Cengage Learning. <https://doi.org/10.1201/b21879-2>
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241–258. <https://doi.org/10.1007/s11704-019-8208-z>
- Dotto, A. C., Dalmolin, R. S. D., Grunwald, S., ten Caten, A., & Pereira Filho, W. (2017). Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. *Soil and Tillage Research*, 172(December 2015), 59–68.
<https://doi.org/10.1016/j.still.2017.05.008>
- Ehrenfeld, J. G., Ravit, B., & Elgersma, K. (2005). Feedback in the plant-soil system. *Annual Review of Environment and Resources*, 30, 75–115.
<https://doi.org/10.1146/annurev.energy.30.050504.144212>
- Eiben, A. E. (2015). Introduction to Evolutionary Computing. In *Kybernetes* (Vol. 33).
<https://doi.org/10.1108/03684920410699216>
- El-Zeiny, M. B., Zawbaa, H. M., & Serag, A. (2021). An evaluation of different bio-inspired feature selection techniques on multivariate calibration models in spectroscopy. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 246, 119042.
<https://doi.org/10.1016/j.saa.2020.119042>

- Fei, Q., Li, M., Wang, B., Huan, Y., Feng, G., & Ren, Y. (2009). Analysis of cefalexin with NIR spectrometry coupled to artificial neural networks with modified genetic algorithm for wavelength selection. *Chemometrics and Intelligent Laboratory Systems*, 97(2), 127–131. <https://doi.org/10.1016/j.chemolab.2009.03.003>
- Gholizadeh, A., Saberioon, M., Ben-Dor, E., Viscarra Rossel, R. A., & Borůvka, L. (2020). Modelling potentially toxic elements in forest soils with vis–NIR spectra and learning algorithms. *Environmental Pollution*, 267. <https://doi.org/10.1016/j.envpol.2020.115574>
- Gobrecht, A., Roger, J. M., & Bellon-Maurel, V. (2014). Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation. A Review. In *Advances in Agronomy* (Vol. 123). <https://doi.org/10.1016/B978-0-12-420225-2.00004-2>
- Gondal, A. H., Hussain, I., Ijaz, A. bakar, Zafar, A., Ch, B. I., Zafar, H., Sohail, M. D., Niazi, H., Touseed, M., Khan, A. A., Tariq, M., Yousuf, H., & Usama, M. (2021). Influence of soil pH and microbes on mineral solubility and plant nutrition: A review . *International Journal of Agriculture and Biological Sciences*, 5(1), 2–12. https://d1wqtxts1xzle7.cloudfront.net/66113631/Volume_5_Issue_1_Paper_8-with-cover-page-v2.pdf?Expires=1639806217&Signature=FRqt9IzQuEtA1LG0eYffRC1830B80os5kz3-tDqJEqNJJWAdYnhZ8L00~uKz-5hrWLUZmCUQgimABd9LjVhOc1X44eq-W-rC8HtbTBEQeJWWTwNzo0D97Di4NnH1DQcBNJY
- Gonzalez, A., & Vargas, N. (2014). Análisis de imágenes hiperespectrales To cite this version : HAL Id : hal-00935014. *Hal*, 9, 14–17. <https://hal.archives-ouvertes.fr/hal-00935014/document>
- González, M., & Montaña, L. (2015). La espectroscopia y su tecnología: Un repaso histórico y su importancia para el siglo XXI. *Latin-American Journal of Physics Education*, 9(4), 13.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144. <https://doi.org/10.1145/3422622>
- Guiñón, J. L., Ortega, E., García-Antón, J., & Pérez-herranz, V. (2007). Moving Average and Savitzki-Golay Smoothing Filters Using Mathcad. *International Conference on Engineering Education*, 1, 1–4. <http://academic.research.microsoft.com/Paper/12119855.aspx>
- Guo, P., Li, T., Gao, H., Chen, X., Cui, Y., & Huang, Y. (2021). Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using vis-nir spectroscopy. *Remote Sensing*, 13(19), 1–20. <https://doi.org/10.3390/rs13194000>
- Hastie, T., Tibshirani, R., & Friedman, J. (2005). The Elements of Statistical Learning : Data Mining , Inference and Prediction Probability Theory : The Logic of Science The Fundamentals of Risk Measurement Mathematicians , pure and applied , think there is something weirdly different about. *The Mathematical Intelligencer*, 27(2), 83–85. <http://link.springer.com/article/10.1007/BF02985802?LI=true#>
- He, Y., Huang, M., García, A., Hernández, A., & Song, H. (2007). Prediction of soil macronutrients content using near-infrared spectroscopy. *Computers and Electronics in Agriculture*, 58(2), 144–153. <https://doi.org/10.1016/j.compag.2007.03.011>
- He, Y., Xiao, S., Nie, P., Dong, T., Qu, F., & Lin, L. (2017). Research on the Optimum Water Content of Detecting Soil Nitrogen Using Near Infrared Sensor. *Sensors (Switzerland)*, 17(2045), 1–12. <https://doi.org/10.3390/s17092045>
- Helland, I. S. (1990). Partial Least Squares Regression and Statistical Models. *Scandinavian Journal of Statistics*, 17(2), 97–114. <http://www.jstor.org/stable/4616159>
- Hollas, J. M. (2011). Modern Spectroscopy. In *Voенно-медицинский журнал* (Vol. 332, Issue 2).
- Hong, Y., Chen, Y., Yu, L., Liu, Y., Liu, Y., Zhang, Y., Liu, Y., & Cheng, H. (2018). Combining

- fractional order derivative and spectral variable selection for organic matter estimation of homogeneous soil samples by VIS-NIR spectroscopy. *Remote Sensing*, 10(3).
<https://doi.org/10.3390/rs10030479>
- Horta, M. D. C., & Torrent, J. (2007). The Olsen P method as an agronomic and environmental test for predicting phosphate release from acid soils. *Nutrient Cycling in Agroecosystems*, 77(3), 283–292. <https://doi.org/10.1007/s10705-006-9066-2>
- Huang, G., Zhu, Q., & Siew, C. (2006). Extreme learning machine : Theory and applications. *Neurocomputing*, 70, 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
- Inoue, Y. (2020a). Satellite- and drone-based remote sensing of crops and soils for smart farming—a review. *Soil Science and Plant Nutrition*, 66(6), 798–810.
<https://doi.org/10.1080/00380768.2020.1738899>
- Inoue, Y. (2020b). Soil Science and Plant Nutrition Satellite- and drone-based remote sensing of crops and soils for smart farming – a review. *Soil Science and Plant Nutrition*, 66(6), 798–810. <https://doi.org/10.1080/00380768.2020.1738899>
- Isaac, W., & Na, A. (2017). On-the-go soil nitrogen sensor based on near infrared spectroscopy. *2016 International Conference on Information Technology, IncITE 2016 - The Next Generation IT Summit on the Theme - Internet of Things: Connect Your Worlds*, 312–315.
<https://doi.org/10.1109/INCITE.2016.7857637>
- Isaak, S., Yusof, Y., Ngajikin, N. H., Ramli, N., & Wen, C. M. (2019). A low cost spectroscopy with Raspberry Pi for soil macronutrient monitoring. *Telkomnika (Telecommunication Computing Electronics and Control)*, 17(4), 1867–1875.
<https://doi.org/10.12928/TELKOMNIKA.V17I4.12775>
- Ji-yong, S., Xiao-bo, Z., Xiao-wei, H., Jie-wen, Z., Yanxiao, L., Limin, H., & Jianchun, Z. (2013). *Rapid detecting total acid content and classifying different types of vinegar based on near infrared spectroscopy and least-squares support vector machine*. 138, 192–199.
<https://doi.org/10.1016/j.foodchem.2012.10.060>
- Ji, W., Adamchuk, V. I., Biswas, A., Dhawale, N. M., Sudarsan, B., Zhang, Y., Viscarra Rossel, R. A., & Shi, Z. (2016). Assessment of soil properties in situ using a prototype portable MIR spectrometer in two agricultural fields. *Biosystems Engineering*, 152, 14–27.
<https://doi.org/10.1016/j.biosystemseng.2016.06.005>
- Ji, W., Li, S., Chen, S., Shi, Z., Viscarra Rossel, R. A., & Mouazen, A. M. (2016). Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. *Soil and Tillage Research*, 155, 492–500.
<https://doi.org/10.1016/j.still.2015.06.004>
- Jiang, C., Zhao, J., Ding, Y., & Li, G. (2023). Vis-NIR Spectroscopy Combined with GAN Data Augmentation for Predicting Soil Nutrients in Degraded Alpine Meadows on the Qinghai-Tibet Plateau. *Sensors (Basel, Switzerland)*, 23(7). <https://doi.org/10.3390/s23073686>
- Jin, J., Chen, Z., Li, L., Steponavicius, R., Thennadil, S. N., Yang, J., & Yu, R. (2012). Quantitative Spectroscopic Analysis of Heterogeneous Mixtures : The Properties of Samples. *Analytical Chemistry*, 320–326.
- Kang, H., Gao, H., & Yu, W. (2017). Evaluation of Spectral Pretreatments , Spectral Range and Regression Methods for Quantitative Spectroscopic Analysis of Soil Organic Carbon Composition. *Spectroscopy Letters*, 7010(March).
<https://doi.org/10.1080/00387010.2017.1297956>
- Kawamura, K., Nishigaki, T., Tsujimoto, Y., Andriamananjara, A., Rabenaribo, M., Asai, H., Rakotoson, T., & Razafimbelo, T. (2021). Exploring relevant wavelength regions for estimating soil total carbon contents of rice fields in Madagascar from Vis-NIR spectra with sequential application of backward interval PLS. *Plant Production Science*, 24(1), 1–

14. <https://doi.org/10.1080/1343943X.2020.1785898>
- Kawamura, K., Tsujimoto, Y., Rabenarivo, M., Asai, H., Andriamananjara, A., & Rakotoson, T. (2017). Vis-NIR spectroscopy and PLS regression with waveband selection for estimating the total C and N of paddy soils in Madagascar. *Remote Sensing*, 9(10). <https://doi.org/10.3390/rs9101081>
- Keyvan, K., Sohrabi, M. R., & Motiee, F. (2021). An intelligent method based on feed-forward artificial neural network and least square support vector machine for the simultaneous spectrophotometric estimation of anti hepatitis C virus drugs in pharmaceutical formulation and biological fluid. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 263, 120190. <https://doi.org/10.1016/j.saa.2021.120190>
- Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., & Abbas, A. (2018). Modern Trends in Hyperspectral Image Analysis: A Review. *IEEE Access*, 6(c), 14118–14129. <https://doi.org/10.1109/ACCESS.2018.2812999>
- Kim, H. J., Sudduth, K. A., & Hummel, J. W. (2009). Soil macronutrient sensing for precision agriculture. *Journal of Environmental Monitoring*, 11(10), 1810–1824. <https://doi.org/10.1039/b906634a>
- Kose, U., Prasath, V. B. S., Mondal, M. R. H., Podder, P., & Bharati, S. (2022). Artificial Intelligence and Smart Agriculture Applications. In *Artificial Intelligence and Smart Agriculture Applications*. <https://doi.org/10.1201/9781003311782>
- Laskar, S., & Mukherjee, S. (2016). Optical Sensing Methods for Assessment of Soil Macro-nutrients and other Properties for Application in Precision Agriculture: A review. *ADB- Journal of Engineering Technology AJET*, 4(1), 206.
- Legnaioli, S., Lorenzetti, G., Pardini, L., Cavalcanti, G. H., & Palleschi, V. (2014). Double and multiple pulse LIBS techniques. In *Laser-Induced Breakdown Spectroscopy* (Vol. 182, pp. 117–141). https://doi.org/10.1007/978-3-642-45085-3_5
- Lei, T., & Sun, D. W. (2022). Achieving joint calibration of soil Vis-NIR spectra across instruments, soil types and properties by an attention-based spectra encoding-spectra/property decoding architecture. *Geoderma*, 405(May 2021), 115449. <https://doi.org/10.1016/j.geoderma.2021.115449>
- Lei, Z., Yao, M., Liu, M., Li, Q., & Mao, H. (2011). Comparison between fertilization N, P, K and No fertilization N, P, K in paddy soil by laser induced breakdown spectroscopy. *Proceedings - 4th International Conference on Intelligent Computation Technology and Automation, ICICTA 2011*, 1, 363–366. <https://doi.org/10.1109/ICICTA.2011.102>
- Li, H., Jia, S., & Le, Z. (2019). Quantitative Analysis of Soil Total Nitrogen Using Hyperspectral Imaging Technology with Extreme Learning Machine. *Sensors (Switzerland)*.
- Li, Hong, Wang, J., Zhang, J., Liu, T., Acquah, G. E., & Yuan, H. (2022). Combining Variable Selection and Multiple Linear Regression for Soil Organic Matter and Total Nitrogen Estimation by DRIFT-MIR Spectroscopy. *Agronomy*, 12(3). <https://doi.org/10.3390/agronomy12030638>
- Li, Y., Yang, Q., Chen, M., Wang, M., & Zhang, M. (2019). An ISE-based On-Site Soil Nitrate Nitrogen Detection System. *Sensors (Switzerland)*, 3.
- Liakos, K. G., Busato, P., Moshou, D., & Pearson, S. (2018). Machine Learning in Agriculture : A Review. *Sensors (Basel, Switzerland)*, 18(1), 1–29. <https://doi.org/10.3390/s18082674>
- Lin, Z. D., Wang, Y. B., Wang, R. J., Wang, L. S., Lu, C. P., Zhang, Z. Y., Song, L. T., & Liu, Y. (2017). Improvements of the Vis-NIRS Model in the Prediction of Soil Organic Matter Content Using Spectral Pretreatments, Sample Selection, and Wavelength Optimization. *Journal of Applied Spectroscopy*, 84(3), 529–534. <https://doi.org/10.1007/s10812-017-0505-4>
- Linker, R., Shmulevich, I., Kenny, A., & Shaviv, A. (2005). Soil identification and chemometrics

- for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy. *Chemosphere*, 61(5), 652–658. <https://doi.org/10.1016/j.chemosphere.2005.03.034>
- Liu, J., Xie, J., Han, J., Wang, H., Sun, J., Li, R., & Li, S. (2020). Visible and near-infrared spectroscopy with chemometrics are able to predict soil physical and chemical properties. *Journal of Soils and Sediments*, 20(7), 2749–2760. <https://doi.org/10.1007/s11368-020-02623-1>
- Liu, Lanfa, Ji, M., & Buchroithner, M. (2018). Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. *Sensors (Switzerland)*, 18(9). <https://doi.org/10.3390/s18093169>
- Liu, Lijuan, Shen, B., & Wang, X. (2014). Research on kernel function of support vector machine. *Lecture Notes in Electrical Engineering*, 260 LNEE, 827–834. https://doi.org/10.1007/978-94-007-7262-5_93
- Lourembam, D., Laskar, S., & Mukherjee, S. (2017). Framework for an optical sensor system for monitoring of soil nitrogen and tailoring soil pH. *Journal of Optics*. <https://doi.org/10.1007/s12596-017-0434-x>
- Lv, N., Ma, H., Member, S., Chen, C., & Member, S. (2021). Remote Sensing Data Augmentation Through Adversarial Training. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 9318–9333. <https://doi.org/10.1109/JSTARS.2021.3110842>
- Lyon, L. A., Keating, C. D., Fox, A. P., Baker, B. E., He, L., Nicewarner, S. R., Mulvaney, S. P., & Natan, M. J. (1998). Raman Spectroscopy. *Analytical Chemistry*, 70(12), 341–361.
- Ma, F., Du, C. W., Zhou, J. M., & Shen, Y. Z. (2019). Investigation of soil properties using different techniques of mid-infrared spectroscopy. *European Journal of Soil Science*, 70(1), 96–106. <https://doi.org/10.1111/ejss.12741>
- Ma, J., Cheng, J., Wang, J., Pan, R., He, F., Yan, L., & Xiao, J. (2022). Rapid detection of total nitrogen content in soil based on hyperspectral technology. *Information Processing in Agriculture*, 9(4), 566–574. <https://doi.org/10.1016/j.inpa.2021.06.005>
- MacAbiog, R. E. N., Fadchar, N. A., & Cruz, J. C. D. (2020). Soil NPK Levels Characterization Using Near Infrared and Artificial Neural Network. *Proceedings - 2020 16th IEEE International Colloquium on Signal Processing and Its Applications, CSPA 2020, Cspa*, 141–145. <https://doi.org/10.1109/CSPA48992.2020.9068717>
- Mammadov, E., Denk, M., Riedel, F., Ka, C., Lewinska, K., Łukowiak, R., Grzebisz, W., & Mamedov, A. I. (2022). *Determination of Mehlich 3 Extractable Elements with Visible and Near Infrared Spectroscopy in a Mountainous Agricultural Land , the Caucasus Mountains*.
- Manolakis, D., Lockwood, R., & Cooley, T. (2016). Hyperspectral Imaging Remote Sensing_Physics, Sensors, and Algorithms. In *Hyperspectral Imaging Remote Sensing*.
- Masrie, M., Rosli, A. Z. M., Sam, R., Janin, Z., & Nordin, M. K. (2019). Integrated optical sensor for NPK Nutrient of Soil detection. *2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application, ICSIMA 2018, November*, 28–30. <https://doi.org/10.1109/ICSIMA.2018.8688794>
- Masrie, M., Rosman, M. S. A., Sam, R., & Janin, Z. (2018). Detection of nitrogen, phosphorus, and potassium (NPK) nutrients of soil using optical transducer. *2017 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2017, 2017-Novem(November)*, 1–4. <https://doi.org/10.1109/ICSIMA.2017.8312001>
- Mendes, W. de S., Sommer, M., Koszinski, S., & Wehrhan, M. (2021). Local Peatlands Spectral Data Influence in Global Spectral Modelling of Soil Organic Carbon and Total Nitrogen Using Near-Infrared Spectrum. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3983929>

- Min, M., & Lee, W. S. (2005). Determination of Significant Wavelengths and Prediction of Nitrogen Content for Citrus. *American Society of Agricultural Engineers ISSN*, 48(1998), 455–461.
- Mirzaee, S., Ghorbani-Dashtaki, S., Mohammadi, J., Asadi, H., & Asadzadeh, F. (2016). Spatial variability of soil organic matter using remote sensing data. *Catena*, 145, 118–127. <https://doi.org/10.1016/j.catena.2016.05.023>
- Mohamed, E. S., El Baroudy, A. A., El-beshbeshy, T., Emam, M., Belal, A. A., Elfadaly, A., Aldosari, A. A., Ali, A. M., & Lasaponara, R. (2020). Vis-nir spectroscopy and satellite landsat-8 oli data to map soil nutrients in arid conditions: A case study of the northwest coast of egypt. *Remote Sensing*, 12(22), 1–20. <https://doi.org/10.3390/rs12223716>
- Montesinos-López, O. A., Montesinos-López, A., & Corssa, J. (2022). Over fitting , Model Tuning , and Evaluation of Prediction Performance. In F. van Euwijk (Ed.), *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 109–139). Springer.
- Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Wiebensohn, J., Bill, R., & Mouazen, A. M. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering*, 152, 104–116. <https://doi.org/10.1016/j.biosystemseng.2016.04.018>
- Moros Portolés, J. (2007). *TRATAMIENTO NUMÉRICO DE LOS DATOS EN EL ANÁLISIS CUANTITATIVO POR ESPECTROMETRÍA VIBRACIONAL* (Vol. 3, Issue September). Universitat de Valencia.
- Mukherjee, S., & Laskar, S. (2019). Vis–NIR-based optical sensor system for estimation of primary nutrients in soil. *Journal of Optics (India)*, 48(1), 87–103. <https://doi.org/10.1007/s12596-019-00517-1>
- Munnaf, M. A., Guerrero, A., Nawar, S., Haesaert, G., Van Meirvenne, M., & Mouazen, A. M. (2021). A combined data mining approach for on-line prediction of key soil quality indicators by Vis-NIR spectroscopy. *Soil and Tillage Research*, 205(June 2020), 104808. <https://doi.org/10.1016/j.still.2020.104808>
- Murphy, K. P. (2012). Machine Learning A Probabilistic Perspective. In *Chance Encounters: Probability in Education*. https://doi.org/10.1007/978-94-011-3532-0_2
- Nawar, S., Delbecq, N., Declercq, Y., De Smedt, P., Finke, P., Verdoodt, A., Van Meirvenne, M., & Mouazen, A. M. (2019). Can spectral analyses improve measurement of key soil fertility parameters with X-ray fluorescence spectrometry? *Geoderma*, 350(May), 29–39. <https://doi.org/10.1016/j.geoderma.2019.05.002>
- Nie, P. C., Dong, T., He, Y., & Qu, F. (2017). Detection of soil nitrogen using near infrared sensors based on soil pretreatment and algorithms. *Sensors (Switzerland)*, 17(5), 1–13. <https://doi.org/10.3390/s17051102>
- Nie, P., Dong, T., He, Y., & Xiao, S. (2018). Research on the Effects of Drying Temperature on Nitrogen Detection of Different Soil Types by Near. *Sensors (Switzerland)*. <https://doi.org/10.3390/s18020391>
- Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B., & Montanarella, L. (2014). Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biology and Biochemistry*, 68, 337–347. <https://doi.org/10.1016/j.soilbio.2013.10.022>
- Norris, K. H., & Williams, P. C. (1984). Optimization of Mathematical Treatments of Raw Near-Infrared Signal in the Measurement of Protein in Hard Red Spring Wheat. *Cereal Chemistry*, 2, 158–165.
- Papousek, D., & Aliev, M. R. (1982). *Molecular Vibration-Rotation Spectra* (R. Zahradnik (ed.);

- 1st ed.). Elsevier.
- Patel, A. K., Ghosh, J. K., & Sayyad, S. U. (2020). Fractional abundances study of macronutrients in soil using hyperspectral remote sensing. *Geocarto International*, 0(0), 1–20. <https://doi.org/10.1080/10106049.2020.1720315>
- Petropoulos, G. P., Arvanitis, K., & Sigrimis, N. (2012). Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Expert Systems with Applications*, 39(3), 3800–3809. <https://doi.org/10.1016/j.eswa.2011.09.083>
- Qi, H., Paz-Kagan, T., Karnieli, A., Jin, X., & Li, S. (2018). Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. *Soil and Tillage Research*, 175(October 2017), 267–275. <https://doi.org/10.1016/j.still.2017.09.006>
- Reeves, J. B. (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*, 158(1–2), 3–14. <https://doi.org/10.1016/j.geoderma.2009.04.005>
- 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-Pérez, J. R., Marcelo, V., Pereira-Obaya, D., García-Fernández, M., & Sanz-Ablanedo, E. (2021). Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards. *Agronomy*, 11(10). <https://doi.org/10.3390/agronomy11101895>
- Rojas, R. (1996). The Backpropagation Algorithm. In *Neural Networks* (pp. 150–182). Springer Berlin Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-61068-4_7
- Sadgrove, E. J., Falzon, G., Miron, D., & Lamb, D. W. (2021). The segmented colour feature extreme learning machine: Applications in agricultural robotics. *Agronomy*, 11(11), 1–16. <https://doi.org/10.3390/agronomy11112290>
- Sáez-Plaza, P., García, A., & Martín, J. (2019). An annotation on the Kjeldahl method. *Anales de La Real Academia Nacional de Farmacia*, 85(1), 14–19.
- Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2018). Recent Advances in Recurrent Neural Networks. *Neural and Evolutionary Computing*, 1–21.
- Sarathjith, M. C., Das, B. S., Wani, S. P., & Sahrawat, K. L. (2016). Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma*, 267, 1–9. <https://doi.org/10.1016/j.geoderma.2015.12.031>
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627–1639.
- Norma Oficial Mexicana NOM-021-RECNAT-2000, Diario Oficial de la Federación (2002). http://dof.gob.mx/nota_detalle.php?codigo=717582&fecha=31/12/2002
- Shao, Y., & He, Y. (2011). Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy. *Soil Research*, 49(2), 166–172. <https://doi.org/10.1071/SR10098>
- Shepherd, K. D., & Walsh, M. G. (2002). Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Science Society of America Journal*, 66(3), 988–998. <https://doi.org/10.2136/sssaj2002.9880>
- Shi, P., Castaldi, F., van Wesemael, B., & Van Oost, K. (2020). Vis-NIR spectroscopic assessment of soil aggregate stability and aggregate size distribution in the Belgian Loam Belt. *Geoderma*, 357(May 2019), 113958. <https://doi.org/10.1016/j.geoderma.2019.113958>
- Singh, R. (2002). C. V. Raman and the Discovery of the Raman Effect. *Physics in Perspective*, 4(4), 399–420. <https://doi.org/10.1007/s000160200002>
- Sleep, B., Mason, S., Janik, L., & Mosley, L. (2022). Application of visible near-infrared absorbance spectroscopy for the determination of Soil pH and liming requirements for

- broad-acre agriculture. *Precision Agriculture*, 23(1), 194–218.
<https://doi.org/10.1007/s11119-021-09834-7>
- Song, X., Gao, Y., Liu, Z., Zhang, M., & Wan, Y. (2019). Development of a predictive tool for rapid assessment of soil total nitrogen in wheat-corn double cropping system with hyperspectral data. *Environmental Pollutants and Bioavailability*, 5940.
<https://doi.org/10.1080/26395940.2019.1679041>
- 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>
- Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE*, 8(6). <https://doi.org/10.1371/journal.pone.0066409>
- Stevens, A., & Ramirez-Lopez, L. (2013). An Introduction to the prospectr package. In *2013 R Package* (0.1; Vol. 136, Issue 12, pp. 1628-a-1629).
<https://doi.org/10.1176/ajp.136.12.1628-a>
- Sun, J., Shi, S., Yang, J., Gong, W., Qiu, F., Wang, L., Du, L., & Chen, B. (2019). Wavelength selection of the multispectral lidar system for estimating leaf chlorophyll and water contents through the PROSPECT model. *Agricultural and Forest Meteorology*, 266–267(November 2018), 43–52. <https://doi.org/10.1016/j.agrformet.2018.11.035>
- Sutton, R. S., & Barto, A. G. (2019). Reinforcement Learning, An introduction. In MIT (Ed.), 余东华 张鑫宇 孙婷 (2nd ed.). Westchester Publishing Services.
- Thenkabail, P. S., Lyon, J. G., & Huete, A. (2018). HYPERSPECTRAL REMOTE SENSING OF VEGETATION. In P. S. Thenkabail, G. J. Lyon, & A. Huete (Eds.), *Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation* (2nd ed.). CRC Press.
<https://doi.org/10.1201/9781315164151>
- Thurnhofer-hemsi, K., López-rubio, E., & Molina-cabello, M. A. (2020). RADIAL BASIS FUNCTION KERNEL OPTIMIZATION FOR SUPPORT VECTOR MACHINE CLASSIFIERS. *IEEE Transactions on Neural Networks and Learning Systems*.
<https://doi.org/https://doi.org/10.48550/arXiv.2007.08233>
- Tümsavaş, Z. (2017). Application of visible and near infrared reflectance spectroscopy to predict total nitrogen in soil. *Environmental Biology*, 38(September), 409–417.
[https://doi.org/10.22438/jeb/38/5\(SI\)/GM-047](https://doi.org/10.22438/jeb/38/5(SI)/GM-047)
- Tziolas, N., Tsakiridis, N., Ogen, Y., Kalopesa, E., Ben-Dor, E., Theocharis, J., & Zalidis, G. (2020). An integrated methodology using open soil spectral libraries and Earth Observation data for soil organic carbon estimations in support of soil-related SDGs. *Remote Sensing of Environment*, 244(January), 111793. <https://doi.org/10.1016/j.rse.2020.111793>
- Uhlenbeck, G. E., & Goudsmit, S. (1976). Spinning Electrons and the Structure of Spectra. *Nature*, 2(4), 391. [https://doi.org/10.1016/0304-3762\(76\)90072-9](https://doi.org/10.1016/0304-3762(76)90072-9)
- United States Department of Agriculture. (1987). USDA Textural Soil Classification. In *Soil Mechanics Level I Module 3 - USDA Textural Soil Classification* (pp. 1–53).
- Van Reeuwijk, L. P. (2002). *Procedures For Soil Analysis* (L. P. Van Reeuwijk (ed.); 6th ed.). International Soil Reference and Information Centre.
- Vanderlinde, J., & Neuenschwander, D. E. (1994). Classical Electromagnetic Theory. In *American Journal of Physics* (Vol. 62, Issue 7). <https://doi.org/10.1119/1.17492>
- Vapnik, V. (1998). *of Function Estimation*. 55–85. https://doi.org/10.1007/978-1-4615-5703-6_
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural*

- Networks*, 10(5), 988–999. <https://doi.org/10.1109/72.788640>
- Vestergaard, R.-J., Adamchuk, V., & Biswas, A. (2021). Evaluation of Optimized Preprocessing and Modeling Algorithms for Prediction of Soil Properties Using VIS-NIR Spectroscopy. *Sensors (Basel, Switzerland)*. <https://doi.org/doi.org/10.3390/s21206745>
- Vestergaard, R., Vasava, H., Aspinall, D., Chen, S., Gillespie, A., Adamchuk, V., & Biswas, A. (2021). *Evaluation of Optimized Preprocessing and Modeling*. 3.
- Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., ... Ji, W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155, 198–230. <https://doi.org/10.1016/j.earscirev.2016.01.012>
- Viscarra Rossel, R. A., Rizzo, R., Demattê, J. A. M., & Behrens, T. (2010). Spatial Modeling of a Soil Fertility Index using Visible-Near-Infrared Spectra and Terrain Attributes. *Soil Science Society of America Journal*, 74(4), 1293–1300. <https://doi.org/10.2136/sssaj2009.0130>
- 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(1), 88–96. <https://doi.org/10.1016/j.geoderma.2014.01.013>
- Wambugu, N., Chen, Y., Xiao, Z., Tan, K., Wei, M., Liu, X., & Li, J. (2021). Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks : A review. *International Journal of Applied Earth Observations and Geoinformation*, 105, 102603. <https://doi.org/10.1016/j.jag.2021.102603>
- Wan, M., Qu, M., Hu, W., Li, W., Zhang, C., Cheng, H., & Huang, B. (2019). Estimation of soil pH using PXRF spectrometry and Vis-NIR spectroscopy for rapid environmental risk assessment of soil heavy metals. *Process Safety and Environmental Protection*, 132, 73–81. <https://doi.org/10.1016/j.psep.2019.09.025>
- Wang, C., Liu, B., Liu, L., Zhu, Y., Hou, J., Liu, P., & Li, X. (2021). A review of deep learning used in the hyperspectral image analysis for agriculture. In *Artificial Intelligence Review* (Vol. 54, Issue 7). Springer Netherlands. <https://doi.org/10.1007/s10462-021-10018-y>
- Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, 324(1). <https://doi.org/10.1088/1757-899X/324/1/012049>
- Warren, G. P., & Whitehead, D. C. (1988). Available soil nitrogen in relation to fractions of soil nitrogen and other soil properties. *Plant and Soil*, 112(2), 155–165. <https://doi.org/10.1007/BF02139991>
- Wartini, N., Minasny, B., Montazerolghaem, M., & Padian, J. (2019). *Convolutional Neural Network for simultaneous prediction of several soil properties using visible / near-infrared , mid- infrared , and their combined spectra* [University of Sydney]. <https://www.sciencedirect.com/science/article/pii/S0016706119300588>
- Wehrle, R., Welp, G., & Pätzold, S. (2021). Total and hot-water extractable organic carbon and nitrogen in organic soil amendments: Their prediction using portable mid-infrared spectroscopy with support vector machines. *Agronomy*, 11(4). <https://doi.org/10.3390/agronomy11040659>
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236(December 2018), 111402. <https://doi.org/10.1016/j.rse.2019.111402>
- Wenjun, J., Zhou, S., Jingyi, H., & Shuo, L. (2014). In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy. *PLoS ONE*, 9(8). <https://doi.org/10.1371/journal.pone.0105708>

- Wilson, M. J. (2019). The importance of parent material in soil classification: A review in a historical context. *Catena*, 182(May), 104131. <https://doi.org/10.1016/j.catena.2019.104131>
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37–52. https://doi.org/10.1007/978-0-387-87811-9_2
- Wu, M., Wang, S., Pan, S., Terentis, A. C., & Strasswimmer, J. (2021). Deep learning data augmentation for Raman spectroscopy cancer tissue classification. *Scientific Reports*, 1–13. <https://doi.org/10.1038/s41598-021-02687-0>
- Xiao, S., & He, Y. (2019). Application of near-infrared spectroscopy and multiple spectral algorithms to explore the effect of soil particle sizes on soil nitrogen detection. *Molecules*, 24(13). <https://doi.org/10.3390/molecules24132486>
- Xu, G., Li, Z., & Li, P. (2013). Fractal features of soil particle-size distribution and total soil nitrogen distribution in a typical watershed in the source area of the middle Dan River, China. *Catena*, 101, 17–23. <https://doi.org/10.1016/j.catena.2012.09.013>
- Xu, S., Zhao, Y., Wang, M., & Shi, X. (2018). Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. *Geoderma*, 310(July 2017), 29–43. <https://doi.org/10.1016/j.geoderma.2017.09.013>
- Yan, X. T., Donaldson, K. M., Davidson, C. M., & Gao, Y. (2018). Effects of sample pretreatment and particle size on the determination of nitrogen in soil by portable LIBS and potential use on robotic-borne remote Martian and agricultural soil analysis systems. *RSC Advances*, 36886–36894. <https://doi.org/10.1039/C8RA07065B>
- Yang, H., Kuang, B., & Mouazen, A. M. (2012). Quantitative analysis of soil nitrogen and carbon at a farm scale using visible and near infrared spectroscopy coupled with wavelength reduction. *European Journal of Soil Science*, 63(3), 410–420. <https://doi.org/10.1111/j.1365-2389.2012.01443.x>
- Yang, Haiqing. (2011). Spectroscopic calibration for soil N and C measurement at a farm scale. *Procedia Environmental Sciences*, 10(PART A), 672–677. <https://doi.org/10.1016/j.proenv.2011.09.108>
- 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(January), 114616. <https://doi.org/10.1016/j.geoderma.2020.114616>
- Yang, M., Mouazen, A., Zhao, X., & Guo, X. (2020). Assessment of a soil fertility index using visible and near-infrared spectroscopy in the rice paddy region of southern China. *European Journal of Soil Science*, 71(4), 615–626. <https://doi.org/10.1111/ejss.12907>
- Yang, M., Xu, D., Chen, S., Li, H., & Shi, Z. (2019). Evaluation of machine learning approaches to predict soil organic matter and pH using vis-NIR spectra. *Sensors (Switzerland)*, 19(2). <https://doi.org/10.3390/s19020263>
- Yang, X. S. (2009). Firefly algorithms for multimodal optimization. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5792 LNCS, 169–178. https://doi.org/10.1007/978-3-642-04944-6_14
- Yao, Xiangqian, Yang, W., Li, M., Zhou, P., Chen, Y., Hao, Z., & Liu, Z. (2018). Prediction of Total Nitrogen in Soil Based on Random Frog Leaping Wavelet Neural Network. *IFAC-PapersOnLine*, 51(17), 660–665. <https://doi.org/10.1016/j.ifacol.2018.08.121>
- Yao, Xiong, Yu, K., Deng, Y., Liu, J., & Lai, Z. (2019). Spatial variability of soil organic carbon and total nitrogen in the hilly red soil region of Southern China. *Springer*.
- Young, A. T. (1981). Rayleigh scattering. *Applied Optics*, 20(4), 533–535.
- Yuan, J., Wang, X., Yan, C., Chen, S., Wang, S., Zhang, J., Xu, Z., Ju, X., Ding, N., Dong, Y., &

- Zhang, W. (2020). Wavelength selection for estimating soil organic matter contents through the radiative transfer model. *IEEE Access*, 8, 176286–176293. <https://doi.org/10.1109/ACCESS.2020.3026813>
- Yubing Wang, Huang, H., & Chen, X. (2021). Predicting Organic Matter Content, Total Nitrogen and pH Value of Lime Concretion Black Soil Based on Visible and Near Infrared Spectroscopy. *Eurasian Soil Science*, 54(11), 1681–1688. <https://doi.org/10.1134/S1064229321110144>
- Yun, Y. H., Li, H. D., Deng, B. C., & Cao, D. S. (2019). An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, 113, 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>
- Yun, Y. H., Wang, W. T., Tan, M. L., Liang, Y. Z., Li, H. D., Cao, D. S., Lu, H. M., & Xu, Q. S. (2014). A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Analytica Chimica Acta*, 807, 36–43. <https://doi.org/10.1016/j.aca.2013.11.032>
- Yun, Y., Li, H., Deng, B., & Cao, D. (2019). *Trends in Analytical Chemistry An overview of variable selection methods in multivariate analysis of near-infrared spectra*. 113, 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>
- Zhang, Jian, Gao, B., Chai, H., Ma, Z., & Yang, G. (2016). Identification of DNA-binding proteins using multi-features fusion and binary firefly optimization algorithm. *BMC Bioinformatics*, 17(1), 1–12. <https://doi.org/10.1186/s12859-016-1201-8>
- Zhang, Jifeng, Wang, Z., Fan, B., Hou, Y., Dou, Y., Ren, Z., & Chen, X. (2020). *Investigating the Proper Application Rate of Nitrogen under Mulched Drip Irrigation to Improve the Yield and Quality of Tomato in Saline Soil*.
- Zhang, Y., Li, M. Z., Zheng, L. H., Zhao, Y., & Pei, X. (2016). Soil nitrogen content forecasting based on real-time NIR spectroscopy. *Computers and Electronics in Agriculture*, 124, 29–36. <https://doi.org/10.1016/j.compag.2016.03.016>
- Zhang, Y., Li, M., Zheng, L., Qin, Q., & Suk, W. (2019). Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm. *Geoderma*, 333(June 2018), 23–34. <https://doi.org/10.1016/j.geoderma.2018.07.004>
- Zhou, P., Yang, W., Li, M., Yao, X., & Liu, Z. (2018). Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Total Nitrogen Detector Performance Analysis of Vehicle-mounted Soil Tot. *IFAC-PapersOnLine*, 51(17), 51–56. <https://doi.org/10.1016/j.ifacol.2018.08.071>
- Zhou, P., Zhang, Y., Yang, W., Li, M., Liu, Z., & Liu, X. (2019). Development and performance test of an in-situ soil total nitrogen-soil moisture detector based on near-infrared spectroscopy. *Computers and Electronics in Agriculture*, 160(January), 51–58. <https://doi.org/10.1016/j.compag.2019.03.016>