



Universidad Miguel Hernández
Facultad de Ciencias Experimentales
Grado en Ciencias Ambientales

Trabajo Fin de Grado

Determinación de propiedades de suelos de Senegal mediante espectroscopia VNIR y *machine learning*

Septiembre 2023

Código COIR: TFG.GCA.IMP.LMP.230531

Autora

Lucía Martínez Pérez

Directores

Ignacio Meléndez Pastor e Ignacio Gómez Lucas
Departamento de Agroquímica y Medio Ambiente

Determinación de propiedades de suelos de Senegal mediante espectroscopia VNIR y *machine learning*

Resumen

El aumento de la población mundial, especialmente en países en vías de desarrollo como Senegal, genera enormes presiones sobre los recursos naturales. En este sentido, la estructura poblacional de su medio rural y las prácticas agrosilvopastoriles tradicionales, condicionan el manejo y estado de degradación de sus recursos edáficos, que además se ven seriamente amenazados por la expansión del desierto del Sáhara.

Este trabajo emplea espectroscopía VNIR para la caracterización de diversos suelos agrícolas de Senegal, con la finalidad de evaluar su posible uso como una herramienta de relativo bajo coste, para el análisis *in situ* del estado de degradación de suelos agrícolas. Se construirá una librería espectral de los suelos muestreados de Senegal para posteriormente desarrollar modelos matemáticos y aplicar técnicas de *machine learning* que nos permitan predecir propiedades edáficas a partir de sus espectros de reflectancia. Los resultados obtenidos son un punto de inicio para continuar con futuras investigaciones en esta prometedora línea de trabajo.

Palabras clave: Senegal, degradación del suelo, espectroscopía VNIR, *machine learning*.

Determination of soil properties in Senegal using VNIR spectroscopy and machine learning

Abstract

The increase in world population, especially in developing countries like Senegal, puts enormous pressure on natural resources. In this sense, the population structure of its rural environment and traditional agrosilvopastoral practices condition the management and state of degradation of its edaphic resources, which are also seriously threatened by the expansion of the Sahara Desert.

This work uses VNIR spectroscopy for the characterization of various agricultural soils in Senegal, in order to evaluate its possible use as a relatively low-cost tool for in situ analysis of the state of agricultural soil degradation. A spectral library of the sampled soils from Senegal will be built to later develop mathematical models and apply machine learning techniques that allow us to predict soil properties from their reflectance spectra. The results obtained are a starting point to continue with future research in this promising line.

Keywords: Senegal, soil degradation, VNIR spectroscopy, machine learning

Agradecimientos

En primer lugar, agradecer a Ignacio Gómez, Ignacio Meléndez. y José Navarro por haber confiado en mí para hacer este proyecto, a Ana P. por ayudarme y enseñarme tanto en las prácticas, a Eli por todos los momentos de risas en el laboratorio mientras tenía que esperar a que la batería de la cámara se cargase y poder continuar, a Mario, Nicol y Bernat por los momentos en los que me quedaba en blanco en el laboratorio y me ayudaban a continuar.

A mi familia por haberme costeadado los estudios (que ya he terminado y seguro estaréis felices por ello), a todos mis compañeros de la universidad Marta, Vicky, Pablo, Vico, María y a todos los que he dejado en el tintero que pronto nos encontraremos en el mundo laboral y seguro que haremos lo posible para aportar nuestro grano de arena y mejorar todo lo que esté a nuestro alcance.

A mis amigos Cristina, Rose y Charquitos, por aquellos momentos de estrés en los que nos tomábamos algo en el bar. A mi cuñada Lili, que ha sido para mí una correctora de primera.

Y por último a mi pareja por estar ahí siempre conmigo apoyándome a continuar en lo que más me gusta: mi futura profesión.

Índice

1. INTRODUCCIÓN	5
1.1. PROBLEMÁTICA DE DEGRADACIÓN DE SUELOS EN SENEGAL Y UTILIDAD DE LA RADIOMETRÍA VNIR PARA SU CONTROL	8
2. ANTECEDENTES Y OBJETIVOS	9
3. MATERIAL Y MÉTODOS	11
3.1. DESCRIPCIÓN DEL ÁREA DE ESTUDIO.....	11
3.2. MUESTREO Y ANÁLISIS DE SUELO	13
3.3. REPOSITORIO FOTOGRÁFICO DE LAS MUESTRAS DE SUELO.....	16
3.4. RADIOMETRÍA PROXIMAL VNIR.....	17
3.5. ANÁLISIS ESTADÍSTICOS DESCRIPTIVOS	19
3.6. ESTIMACIÓN DE PROPIEDADES EDÁFICAS MEDIANTE ESPECTROSCOPIA VNIR.....	20
3.7. PREDICCIÓN DE NIVELES DE DEGRADACIÓN DEL SUELO MEDIANTE ESPECTROSCOPIA VNIR Y TÉCNICAS DE APRENDIZAJE AUTOMÁTICO (<i>MACHINE LEARNING</i>).....	22
4. RESULTADOS.....	24
4.1. ANÁLISIS DE LAS PROPIEDADES EDÁFICAS DEL ÁREA DE ESTUDIO.....	24
4.2. DETERMINACIÓN DE PROPIEDADES EDÁFICAS MEDIANTE ESPECTROSCOPIA VNIR.....	27
4.3. APLICACIÓN CONJUNTA DE TÉCNICAS DE <i>MACHINE LEARNING</i> Y RADIOMETRÍA VNIR PARA EVALUAR EL NIVEL DE DEGRADACIÓN DE SUELOS.	32
5. CONCLUSIONES Y PROYECCIÓN FUTURA.....	34
6. BIBLIOGRAFÍA	35
7. ANEXOS	40
7.1. LIBRERÍA DE SUELOS	40

1. Introducción

A día de hoy, la población global aumenta de manera gradual y, por ende, esto genera un aumento de la demanda de recursos naturales para el abastecimiento de la población. Se estima que la demanda de una población creciente aumentará un 60 % la presión sobre el suelo para 2050 (FAO, 2015).

Si comparamos países con distinto nivel de desarrollo, como España y Senegal, se observa como el crecimiento demográfico de Senegal permanece estable mientras que el de España ha sufrido fluctuaciones, y aunque Senegal tiene menor número de habitantes tiene mayor crecimiento demográfico (Figura 1).

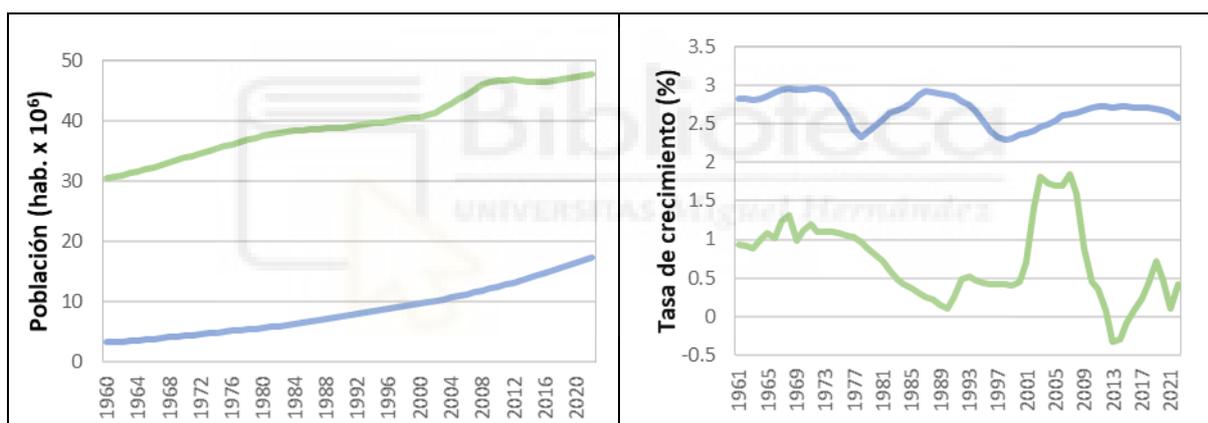


Figura 1. Gráficas comparativas entre España (azul) y Senegal (verde) de las variables: *izquierda*) evolución de la población total (en millones de habitantes) entre 1960 y 2021; y *derecha*) crecimiento demográfico anual (en %) entre 1961 y 2021. Fuente: WorldData Bank, 2023

Según ICEX (2023) y PSE (2023), el principal eje económico de Senegal es gracias al sector agrícola, fundamental para la mayor parte de la población que depende de ellas. De acuerdo con las estadísticas proporcionadas por el Banco Mundial durante el año 2021, el sector agrícola contribuyó con un 15 % al Producto Interno Bruto (PIB) y tuvo una participación del 22 % en la ocupación laboral. Según la FAO, la superficie cultivable en Senegal es el 20% de la extensión del territorio (siendo 3,8 Mha), y de esta se emplean solamente un 65% (Figura 2).

Realizan cultivo de secano y de regadío, el 98% de la superficie cultivada son tierras de secano, y los cultivos de irrigación se ejercen en torno a el valle del río Senegal y la cuenca del Anambé.

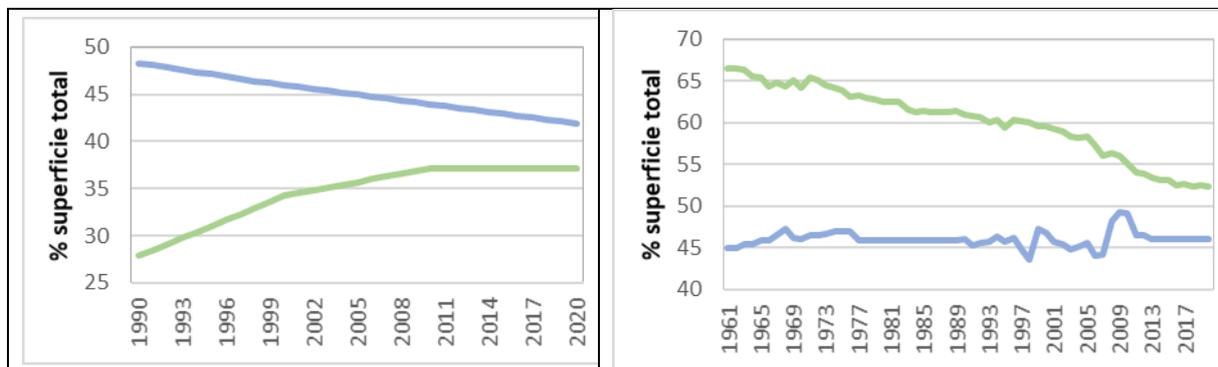


Figura 2. Gráficas comparativas entre España (azul) y Senegal (verde) de las variables: *izquierda*) evolución de la superficie forestal entre 1990 y 2020; y *derecha*) evolución de la superficie de uso agrícola entre 1961 y 2020, ambas en porcentaje de la superficie terrestre total. Fuente: WorldData Bank, 2023

Un 90% de la agricultura es de tipo tradicional, es decir, mano de obra intensiva y poco cualificada, siendo un 7,6% de la productividad total del país. Esta escasa productividad se debe al limitado uso de mecanización, dependiendo así de las importaciones alimentarias para el consumo local.

Gran parte del aprovechamiento agrícola en el país está centrado en el cacahuete, mayormente por empresas industriales, pero también producen azúcar, algodón, aceite, cubitos de caldo, harina y pienso para animales. “La mayor parte de estas empresas se benefician del estatuto de empresa franca” (PSE, 2023). La producción de frutas y hortalizas es mayormente para la demanda local, pero con un potencial uso en exportación de estos productos fuera de temporada ya que se especula que es el sector más beneficioso por su competitividad y productividad.

Entre las variedades de los cultivos de cereales, que sumaron una producción total de 3.640.545 toneladas en el año 2020, se destacan el mijo, el sorgo y el maíz como cultivos propios de terrenos de secano, mientras que el arroz se cultiva con métodos de regadío/irrigación.

Además, es importante resaltar dentro de la diversidad de cultivos presentes en el país, aquellos que requieren una mayor cantidad de recursos agrícolas. Entre estos se encuentran el cacahuete, el arroz, la judía verde y el tomate, los cuales se destinan principalmente a la exportación. Además, otros cultivos como el melón, el pimiento, la patata, la cebolla, la papaya y el mango también ocupan un lugar destacado en la producción agrícola (Figura 3).

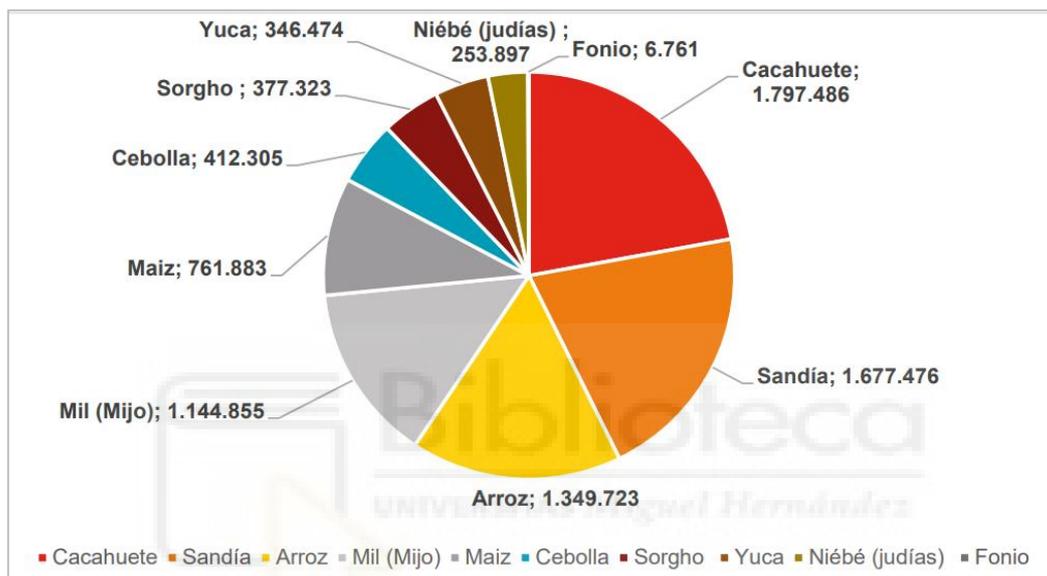


Figura 3. Producción agrícola (en toneladas) de Senegal para el año 2020. Fuente: ICEX (2023).

La importancia del sector agrícola en la economía y prosperidad de la población de Senegal suscita la obvia necesidad de promover prácticas sostenibles de uso del suelo, con la finalidad de minimizar los riesgos de degradación del mismo que, por su ubicación geográfica en la región del Sahel, no está exento de graves amenazas.

1.1. Problemática de degradación de suelos en Senegal y utilidad de la radiometría VNIR para su control

Debido a la sobreexplotación de recursos, actualmente convivimos con la degradación de los suelos, siendo el resultado del empobrecimiento del suelo la reducción de la capacidad de este para abastecer de bienes y servicios ambientales, muchas veces por la disminución de la cobertura vegetal y la reducción en la producción de alimentos. La erosión asociada, provoca un aumento de la sedimentación y eutrofización de los cursos y afecta a las reservas de agua. Frecuentemente, cuando el suelo se degrada, se enmienda las pérdidas productivas con la utilización de fertilizantes (FAO, 2019).

Dada la proximidad de Senegal a la vertiente meridional del desierto del Sáhara, existe un notable riesgo de degradación de los suelos agrícolas que avoque a un proceso de desertificación, si el manejo de los mismos es inadecuado, siendo irreversible e imposible recuperar su estado original. Entre las principales problemáticas de degradación de suelos en Senegal es posible destacar las siguientes (AICD, 2023):

- Salinización de tierras agrícolas y especialmente de arrozales.
- Erosión hídrica que provoca desprendimientos y cárcavas.
- La erosión eólica que elimina la capa superficial de los suelos y destruye su potencial de producción.
- Sobrepastoreo que excede las capacidades de carga.
- Monocultivo de maní.

Estos diferentes fenómenos afectan gravemente las posibilidades de producción agrícola de la población y acentúan su pobreza. Prueba del creciente interés en actual por reducir los riesgos asociados a la degradación de suelos en el área de estudio, son investigaciones como el proyecto LADA o el proyecto DeSurvey (FAO, 2019) para poder valorar la salud de estos suelos en zonas como Senegal. Además, La erosión asociada, provoca un aumento de la sedimentación y eutrofización de los cursos y afecta a las reservas de agua.

Si bien existe un creciente interés por minimizar la degradación de los suelos en países en vías de desarrollo, no es posible omitir la realidad de que son país en los que los recursos

económicos disponibles para dichas actividades son muy escasos, por lo que se requiere de disponer de herramientas adecuadas, que permitan realizar un diagnóstico del estado de degradación de los suelos con un coste reducido y una necesidad de personal técnico limitada. Es en este punto donde se justifica plenamente el disponer de herramientas como la espectroscopía proximal, que permitan realizar un rápido diagnóstico *in situ* del estado de degradación del suelo, permitiendo realizar estimaciones cuantitativas o semicuantitativas de propiedades fisicoquímicas edáficas.

La espectroscopia de reflectancia en el infrarrojo cercano (*Near Infrared Reflectance Spectroscopy*, NIRS) ha alcanzado un gran desarrollo a nivel mundial por su precisión y exactitud (Valenciaga y Saliba, 2006) con un gran potencial para ser automatizada. La espectroscopía infrarroja existe desde hace un siglo, teniendo así una nueva opción a los métodos químicos y químico-biológicos tradicionales. Dentro del ámbito de la observación de espectros vibracionales, los espectrómetros infrarrojos destacan como herramientas de vital importancia. Las características más significativas de esta forma de espectroscopía incluyen lo siguiente: generar la huella digital de una sustancia definida mediante sus espectros infrarrojos, determinar la concentración de una sustancia y analizar muestras siendo una herramienta no destructiva, es una técnica no destructiva, rápida, de gran precisión y exactitud; entre otras propiedades menos relevantes para este estudio (Gómez y Murillo, 2006).

2. Antecedentes y Objetivos

El estudio que vamos a realizar es una continuación de una memoria sobre el medio rural y gestión sostenible de suelos en la Reserva Natural Comunitaria de Dindéfelo (Gallego Moreno, 2019), donde se analizaron muestras de suelos donde garantizar la productividad, riqueza y biodiversidad de la zona. En este, Gallego Moreno plasma la necesidad de comprender y abordar la degradación en aumento de los suelos en Dindéfelo (Senegal) dado al cambio climático, además del auge demográfico y urbano, los cambios dietéticos y degradación edáfica. Por ello, urge garantizar la seguridad alimentaria en un país en vías de desarrollo,

puesto que el aumento de demanda en terrenos cultivables y presiones antrópicas han llevado a la degradación del suelo, afectando a la producción agrícola y salud del ecosistema.

Por ello, Gallego Moreno, refleja la importancia del equilibrio entre desarrollo local, conservación del medio ambiente y gestión sostenible de los suelos, siendo este su principal objetivo junto con la fertilidad de estos en zonas de difícil acceso. Proporcionando a los agricultores y encargados implicados herramientas y métodos prácticos y eficientes para evaluar y potenciar la calidad y salud de los suelos en condiciones de campo, lo cual resulta fundamental en términos de cooperación internacional y el fomento de una agricultura más sostenible.

El objetivo principal a alcanzar es asociar la reflectancia del suelo, mediante el uso de un radiómetro, con las características del suelo analizadas y poder hacer una predicción de esta mediante el espectro obtenido.

Para ello, se procederá a analizar los resultados analíticos y los espectrales mediante diferentes procedimientos estadísticos, que nos permitan extraer información valiosa para la evaluación de la degradación de suelos mediante radiometría proximal. Para tal fin, en el estudio se abarcan una serie de objetivos específicos que planteamos a continuación:

- Analizar desde el punto de vista estadístico, las propiedades edáficas de los suelos de Senegal muestreados, con la finalidad de comprender sus características básicas.
- Crear un repositorio de imágenes a escala microscópica de los suelos muestreados, proporcionando una base de datos visual de los mismos.
- Crear una librería espectral obtenida mediante espectroscopia proximal VNIR, para su posterior empleo en la determinación de propiedades edáficas en base a diferentes procedimientos matemáticos.
- Evaluar el potencial de los espectros VNIR para predicción/estimación de propiedades edáficas de interés, utilizando diversas técnicas de análisis estadísticos y de *machine learning*, con la finalidad de evaluar la potencial aplicabilidad de teles espectros para la determinación in situ del estado de degradación de suelos agrícolas en Senegal.

En conjunto, estos objetivos específicos brindarán una estructura sólida para la investigación, aprovechando métodos y tecnologías avanzadas como los conocimientos previos en el campo para abordar eficazmente la problemática de la degradación del suelo desde diferentes ángulos y con herramientas analíticas innovadoras y contribuir al desarrollo de soluciones sostenibles de manera integral y efectiva.

3. Material y Métodos

Este trabajo combina técnicas analíticas de suelos para la determinación de algunas de sus propiedades fundamentales, con técnicas de espectroscopía VNIR y diversos procedimientos de análisis de datos, utilizando muestras de suelos procedentes de Senegal. A continuación, se presenta la descripción del área de estudio, muy importante para mostrar las características fundamentales del entorno en el que las muestras de suelos fueron recolectadas, así como los diferentes procedimientos de recolección y análisis de suelos, y las diferentes técnicas de procesamiento de datos empleadas.

3.1. Descripción del área de estudio

La República del Senegal tiene una superficie de 196.712 km² y se ubica al oeste de África, haciendo frontera con el océano atlántico y con los países de Mauritania, Mali, Guinea-Conakry, Guinea-Bissau y Gambia (Figura 4). Desde 1960, es independiente de Francia, y se divide en 14 regiones. Su capital es Dakar (localizado al norte) y aunque su idioma oficial es el francés, tiene más de 7 dialectos. En la zona sureste se localiza la región de Kédougou (en la que se ubica la zona de estudio), con 16.800 km² de superficie y más de 150.000 habitantes, siendo una de las regiones con menor densidad de población del país.

El país en cuestión se caracteriza mayormente por ser un territorio de topografía llana, con el Fouta-Djallon al sureste su punto más alto, es decir, su cota máxima, alcanzando los 498 metros de altitud. La línea costera se extiende a lo largo de aproximadamente 600 kilómetros, con un perfil rectilíneo y arenoso en la parte norte, siendo dificultoso para la navegación. No

obstante, la zona central, cuyo origen es volcánico, adopta una forma más curvilínea, brindando la posibilidad de establecer puertos estratégicos, entre los cuales destaca el puerto de Dakar. En cambio, en la zona sur hasta la desembocadura del río Saloum, la costa se torna más pantanosa, donde se despliegan extensas y agradables playas. Por último, en el extremo sur el terreno recupera su complejidad al volverse profundo y tortuoso, dando origen a los estuarios de los Ríos del Sur (ACTSL, 2023).



Figura 4. Mapa de ubicación y principales infraestructuras de comunicación de Senegal. Fuente: ICEX, 2015

Geológicamente, podemos describir a Senegal como una depresión sedimentaria que se fusiona con el zócalo continental en su parte oriental, mientras que en la parte occidental muestra signos de actividad volcánica en las cercanías de la península de Cabo Verde.

Senegal cuenta con la presencia de cinco ríos que fluyen a través de su territorio, siendo el río Senegal el más extenso, con sus 1.600 kilómetros de recorrido que le prestan el nombre a la nación. Los otros ríos, el Faleme, el Gambia, el Casamance y el Saloum. Cada uno de esos ríos tiene un papel único en la hidrología y el paisaje del país. Pueden ofrecer agua para riego,

abastecer a las comunidades con recursos hídricos, influir en la agricultura y la biodiversidad, y en general, formar componentes esenciales del entorno natural y la vida diaria de las personas en Senegal. Es importante destacar que estos ríos siempre se originan en la región de Fouta Djallon, un importante reservorio de agua clave en África occidental (ACTSL, 2023).

La clase climática predominante en el sur de Senegal, que es donde se ubica nuestra zona de estudio, es de tipo tropical sabana (Aw) conforme a la clasificación climática de Köppen (Kottek et al., 2006), caracterizado por presentar un clima cálido durante todo el año y marcado por la existencia de una estación seca acusada. El clima de tipo A es propio de zonas tropicales, con un clima cálido durante todo el año, donde las temperaturas medias superan los 18 °C y las precipitaciones anuales exceden la evaporación. Dentro de esta clase, el tipo w corresponde a la categoría de sabana, en la cual se experimenta un clima cálido durante todo el año, con una estación seca. Este subgrupo se caracteriza por tener meses con precipitaciones inferiores a 60 mm, así como el mes más seco es inferior a:

$$100 - \frac{\text{Precipitación anual}}{25}$$

3.2. Muestreo y análisis de suelo

Las muestras de suelo a analizar proceden de 10 parcelas agrícolas del sur de Senegal recolectadas (Figura 5) en el marco del estudio sobre el “Medio rural y gestión sostenible de suelos en la Reserva Natural Comunitaria de Dindefelo” (Gallego Moreno, C., 2019).

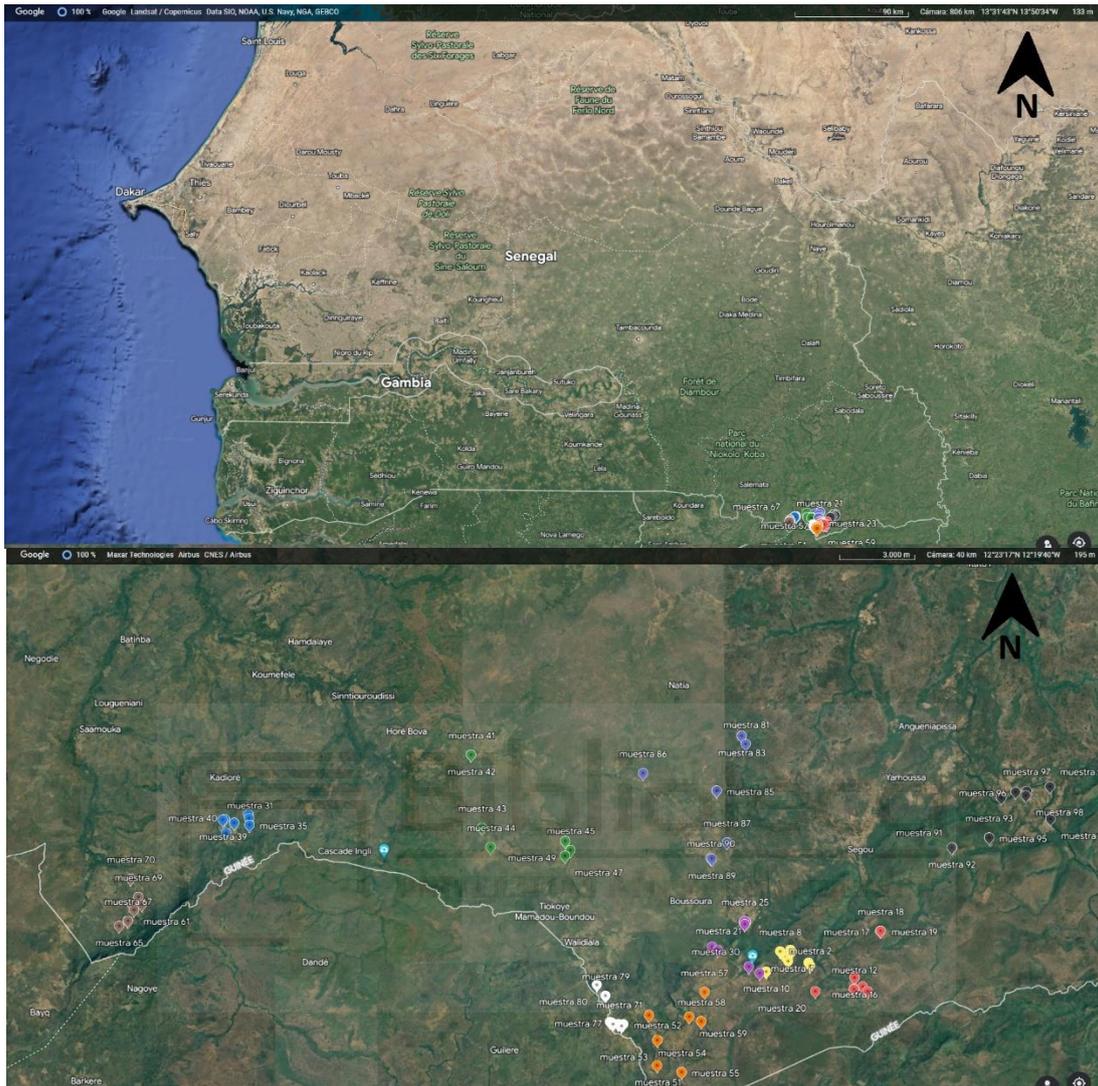


Figura 5. Localización de las muestras de suelos utilizadas en el trabajo. Fuente: Google Earth

Dichas muestras fueron seleccionadas y recolectadas en conjunto por los representantes de Dindéfelo y el instituto Jane Goodall España (IJGE), distribuidas en las inmediaciones de 10 poblaciones del entorno de la reserva natural Dindéfelo (Tabla 1).

Tabla 1. Poblaciones del entorno del de la reserva natural de Dindéfelo en las que se recogieron muestras de suelo. Se incluyen su coordenadas y altitud orientativa.

Nº población	Nombre	Longitud	Latitud	Altitud (m)
1	Afia	12° 22' 09" N	12° 18' 43" W	436
2	Badiari	12° 21' 46" N	12° 17' 05" W	435
3	Dande	12° 21' 44" N	12° 19' 35" W	448
4	Goumbambere	12° 24' 57" N	12° 30' 50" W	124
5	Koukoudji	12° 24' 58" N	12° 24' 42" W	130
6	Nandoumari	12° 20' 44" N	12° 21' 26" W	441
7	Nepen Peuhl	12° 23' 13" N	12° 32' 50" W	154
8	Tanague	12° 20' 47" N	12° 22' 33" W	166
9	Thiangue	12° 24' 31" N	12° 20' 26" W	136
10	Teppere	12° 25' 02" N	12° 13' 40" W	165

En total se recolectaron 100 muestras de suelos independientes, 10 por cada poblado, siendo dichas muestras enviadas a la Estación Biológica Fouta Djallon, sede del IJGE. Allí se dejaron secar y tamizaron a 2 mm, guardándolas posteriormente secas y cerradas para su correcto almacenamiento y evitar posibles alteraciones en su composición, y así poder llegar a las instalaciones de la Universidad Miguel Hernández de Elche (UMH). En la UMH fueron analizadas en laboratorio en dicha campaña de estudio, y ahora en este trabajo, dichas muestras son analizadas mediante radiometría proximal, para analizar las posibilidades de aplicación de la radiometría VNIR en la evaluación del estado de degradación de los suelos. Las propiedades edáficas analizadas en las muestras de estudio fueron las siguientes:

- **Color:** mediante el sistema Munsell (Munsell, 1975) utilizando los parámetros de matiz, intensidad y pureza
- **pH:** mediante un pH-metro de sobremesa (GLP21 Crison) que ayudará a saber qué tipo de cultivo es más idóneo para el suelo.
- **Conductividad eléctrica (CE):** medido con el conductímetro de sobremesa (GLP31 Crison)
- **Textura:** determinando las fracciones arcilla, limo y arena mediante un densímetro Bouyoucos (Gee y Bauder, 1986).

- **Materia orgánica (MOS):** mediante el método de pérdida por ignición (*Loss on Ignition*, LOI) sometiendo la muestra a 380°C por calcinación en un horno-mufla durante 24 horas para lograr la oxidación de materia orgánica (Davies, 1974). Gracias a este tratamiento podremos apreciar la diferencia en los espectros con las muestras después de haber pasado por la mufla y con la muestra que ya teníamos.
- **Nitrógeno, fósforo y potasio:** utilizando la digestión Kjeldahl (Bradstreet, 1954) para la medición de nitrógeno total Kjeldhal (NTK), método Olsen para la medición de fósforo asimilable (Olsen et al., 1954) y por último espectroscopia de emisión con un IPC para el potasio (Houba et al., 2000).

3.3. Repositorio fotográfico de las muestras de suelo

En la primera fase de análisis de este trabajo se procedió a la creación de un repositorio fotográfico de las muestras de suelo disponibles. La cantidad de suelo disponible era bastante escasa, por lo que se procedió utilizaron placas Petri para proceder tanto a la toma de imágenes como a la posterior fase de radiometría VNIR. Para la toma de imágenes, se empleó binocular/microscopio estereoscópico BMS 141 74956 y una cámara fotográfica Canon EOS 450D acoplada en la parte superior del binocular.

Se realizaron cuatro fotografías por cada muestra de suelo, dos de las tomadas directamente del suelo seco y dos tras haber sido sometidas las muestras a la determinación del contenido de materia orgánica (LOI). Además, se realizaron duplicados de las imágenes, incluyendo en la parte superior de la muestra un portaobjetos que lleva impreso un círculo de escala con un diámetro de 3 milímetros de diámetro. De este modo es posible crear un repositorio en el que es posible obtener información del color y escala de partículas (Figura 6), de gran utilidad para poder analizar los resultados de posteriores análisis.



Figura 6. Ejemplo de fotografías de la una muestra de suelo, antes y después de ser sometida a la determinación del contenido en materia orgánica mediante calcinación (LOI). En la fila inferior se muestran las mismas imágenes enfocando el círculo de escala.

3.4. Radiometría proximal VNIR

Una vez tomadas las muestras fotográficas, se creó una librería espectral de las muestras de suelos. Para ello, en una cámara oscura pondremos las muestras en placas Petri, donde no se dejará ninguna zona sin cubrir para evitar la modificación de los espectros por la reflectancia del fondo, bajo el sensor de una fibra óptica iluminada por lámparas halógenas como fuente de iluminación, con rango del espectro visible e infrarrojo bajo y medio en la que trataremos de evitar el riesgo por las sombras creadas por esta y los movimientos con respecto al objetivo del suelo con esta configuración.

Entre la muestra y la fuente de iluminación habrá un ángulo sólido de aproximadamente $22,5^\circ$. Tomaremos mediciones cenitales de las muestras con una empuñadura de pistola dispuesta sobre un trípode a una distancia aproximada de 10 a 30 cm por encima de la muestra con un espectrorradiómetro de superficies Analytical Spectral Devices (ASD) Field Spec Hand Held

VNIR radiometer (ASD Inc., Boulder, CO, USA), posee de amplitud (325-1075) nm rozando el espectro visible e infrarrojo cercano de onda corta (SW-NIR) con una precisión de ± 1 nm y una resolución inferior a 3 nm en el rango de 700 nm (Meléndez-Pastor, et al., 2008). El espectrorradiómetro tendrá conectada una fibra óptica, esta estará conectada a un ordenador configurado especialmente para el programa utilizado. Primeramente, calibramos el sistema y una vez calibrado, tomaremos 5 mediciones radiométricas por cada muestra de suelo configurada a 544 ms. Entre las muestras pondremos una placa de referencia blanca Spectralon 99% para calibrar el equipo y tener la reflectancia base. *“El spectralon o espectralón, es un fluoropolímero que posee la reflexión difusa más alta de cualquier material conocido o capa para las regiones del espectro ultravioleta, visible o infrarrojas, por ello se utiliza para calibrar”* (Forcadell Pérez, G. J., 2022). Estas mediciones de calibración inicial y final se promediaron para obtener un espectro medio del panel de referencia que permitió la corrección de los espectros brutos del suelo a valores de reflectancia.

Cuando obtuvimos los datos de la reflectancia medida con el espectrorradiómetro se importaron los datos a la base de datos del software SAMS. Tras esto, se hizo un preprocesamiento de los espectros para poder apreciar la utilidad de los datos obtenidos, en caso de no ser válidos, se realizaría un cribado de esos datos, haciendo un primer filtro de los datos. Con los datos que pasasen el primer filtro, se realizó un promedio de la serie de espectros obtenidos de cada muestra y a estos nuevos espectros logrados se suavizaron para evitar las fluctuaciones por ruidos, mediante un filtro de ventana móvil Savitzky–Golay con un polinomio de orden 2 y 15 nanómetros de rango a izquierda a derecha (Figura 7). A mayor orden del polinomio, mayor adaptación a la sinuosidad obtendremos, por lo que es preciso que no esté ni muy ajustada la recta, pero tampoco muy recta, solamente se necesita un valor en la recta más suavizado. Cortaremos estos gráficos, para tener solamente el rango desde 400 a 960 nanómetros, que es el rango con espectros más nítidos y libres de ruido.

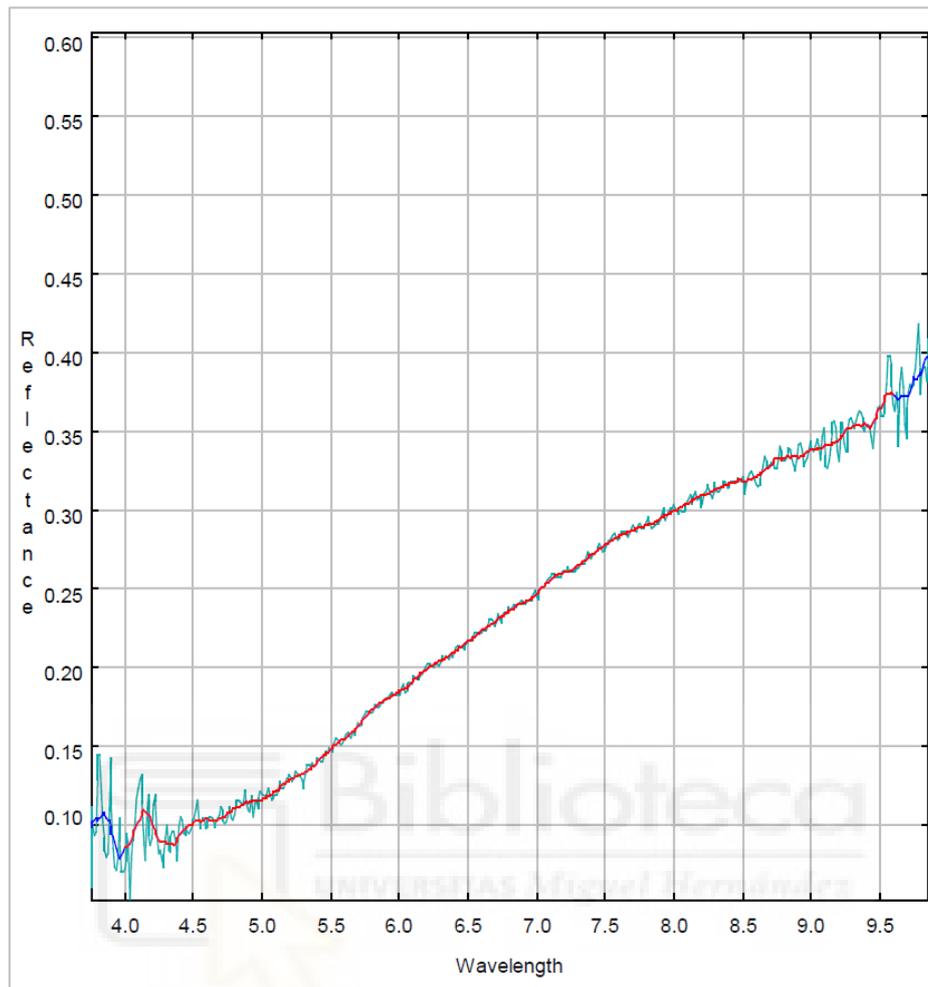


Figura 7. Ejemplo del efecto del suavizado mediante la aplicación del filtro Savitzky–Golay

3.5. Análisis estadísticos descriptivos

Se procedió a realizar un análisis estadístico de las propiedades analíticas de las muestras de suelo empleadas, con la finalidad de conocer sus características básicas. En primer lugar, se analizaron estadísticos descriptivos básicos del contenido en MOS, pH, CE, NTK, P y K., calculando mínimo, máximo, valor promedio y desviación estándar para los suelos de cada población muestreada. Además, se aplicó test estadístico de Shapiro-Wilk para conocer la distribución de datos de las propiedades analíticas y guiarnos en la elección de los siguientes test estadísticos. Dicho test reveló que la distribución predominante de los datos no se ajustaba a una distribución normal y, dado que la transformación de los datos tampoco

permitía obtener una distribución normal, se decidió aplicar métodos no paramétricos para el posterior análisis de los datos.

Se procedió a una comparación de las medias para los valores de las propiedades edáficas, utilizando como factor la población en la que las muestras de suelo fueron recolectadas. El objetivo del análisis era comprender mejor posibles diferencias entre los suelos de las distintas ubicaciones muestreadas. Para ello, se hizo uso del test no paramétrico de Kruskal-Wallis (K-W) haciendo uso del factor población. El resultado de este test es comparable a de una ANOVA de una vía, pero sin la necesidad de asumir una distribución normal de los datos.

Para finalizar este apartado, se procedió a analizar la correlación entre las diferentes variables edáficas, de cara a conocer la relación que presentaban entre sí y posibles autocorrelaciones. Para este procedimiento se utilizó el test rangos de correlación de Spearman, con una conceptualización asimilable al test de correlación de Pearson, pero sin asumir normalidad entre los datos. Adicionalmente, se empleó el método de Holm para determinar la significatividad (p -valor) de las correlaciones de Spearman. Todos estos análisis estadísticos fueron realizados con el lenguaje de programación R (R Core Team, 2023).

3.6. Estimación de propiedades edáficas mediante espectroscopía VNIR

En la siguiente fase del trabajo se procedió a intentar cuantificar las propiedades edáficas utilizando los espectros VNIR obtenidos en el laboratorio. El propósito de este procedimiento es tratar de desarrollar modelos matemáticos que nos permitieran estimar de forma rápida y no destructiva, propiedades del suelo para evaluar su estado de degradación, a partir de las mediciones de radiometría proximal.

Se emplearon regresiones por mínimos cuadrados parciales (PLSR – *Partial Least Square Regression*). El método PLSR ha sido diseñado para enfrentar la situación en la que hay muchas variables predictivas, posiblemente correlacionadas, y relativamente pocas muestras (Mevik y Wehrens, 2007). En este sentido, PLSR proporciona métodos de modelado multivariado cuantitativo viables para quimiometría donde se emplean datos espectrales muy detallados

(es decir, alta resolución espectral o número de bandas) para predecir cuantitativamente un número limitado de muestras problemáticas (Wold et al., 2001).

Para poder realizar la modelización mediante el procedimiento de PLSR, se desarrolló una base de datos que incluía tanto los datos de la librería espectral como los parámetros edáficos de cada suelo muestreado. Seguidamente, se procedió a la división de los datos en dos subconjuntos, uno para la fase de calibración y otro para la fase de validación. Se realizó una división aleatoria estratificada por población en proporción 70:30, lo que implica que el 70% de los suelos de cada población se utilizaban para desarrollar el modelo matemático mediante PLSR y el 30% restante se utilizaba para validar dicho modelo, en aras de analizar su potencial aplicabilidad en muestras independientes a las utilizadas en la fase de entrenamiento del modelo PLSR.

Durante la fase de calibración de los modelos se utilizó validación cruzada *Leave One Out* (LOO), donde se elige una única muestra como grupo de validación y las demás conforman el grupo de entrenamiento. Parámetros tales como el número de componentes de los modelos PLSR, así como sus valores asociados de coeficiente de correlación (R^2), coeficiente de variación (CV) y error cuadrático medio (RMSE -*root mean square error*) fueron obtenidos durante esta fase de calibración y permitieron la selección de los modelos PLSR más idóneos. Seguidamente, dichos modelos fueron validados con el otro subconjunto de muestras independientes, con la finalidad de evaluar la capacidad predictiva de tales modelos. Se evaluó la precisión de tales modelos utilizando parámetros estadísticos tales como el coeficiente de correlación (R^2), coeficiente de variación (CV), error cuadrático medio (RMSE -*root mean square error*), error cuadrático medio normalizado (NRMSE – *normalized root mean square error*), error máximo permitido (MAE – *máximum allowable error*) y desviación relativa permitido (RPD – *relative predictive deviation*). Todas fases de aplicación de las PLSR fueron realizadas con el lenguaje de programación R (R Core Team, 2023).

3.7. Predicción de niveles de degradación del suelo mediante espectroscopía VNIR y técnicas de aprendizaje automático (*Machine Learning*)

En esta última fase del trabajo se emplearon técnicas de aprendizaje automático o *Machine Learning* (ML), con la finalidad de emplear los espectros edáficos de radiometría proximal VNIR, para predecir niveles de degradación de los suelos estudiados. Esto podría proporcionar una herramienta no destructiva y de fácil implementación *in situ*, para analizar la alteración de propiedades edáficas específicas que pudieran relevar problemas de degradación y prácticas agrícolas inadecuadas.

El aprendizaje automático o *Machine Learning* (ML) es una rama dentro de la inteligencia artificial (IA) enfocada en la utilización de datos y algoritmos que simulan el aprendizaje humano con una mejora gradual en precisión. Mediante la aplicación de ML se pueden realizar clasificaciones, predicciones y descubrir conocimientos en proyectos de minería de datos a raíz de datos previos y ejemplos de manera estructurada mediante participación humana, aprendiendo así de la experiencia. Gracias a ello, nos permite identificar patrones, relaciones y tendencias comportamentales en los datos (Witten, Frank y Hall, 2016).

En esta parte del trabajo se empleó el software WEKA (*Waikato Environment for Knowledge Analysis*). Se desarrollaron modelos matemáticos para predecir categorías de contenido en materia orgánica (Tabla 2) y pH (Tabla 3) en los suelos de estudio, en base a los espectros VNIR.

Tabla 2. Clasificación del contenido de materia orgánica para un suelo agrícola. Fuente: Villarroel 1988

Clasificación	% MO
Muy bajo	0,0 – 1,0
Bajo	1,1 – 2,0
Moderado	2,1 – 4,0
Alto	4,1 – 8,0
Muy alto	> 8,0

Se emplearon los procedimientos de *machine learning* árboles de decisión y *random forest* para el desarrollo de los modelos matemáticos. Los árboles de decisión son una técnica de aprendizaje automático que se utiliza para construir evaluar modelos y tomar decisiones basadas en datos de manera sencilla., funcionan bien en el conjunto de entrenamiento, es decir, utilizado para la clasificación y la regresión. Se puede pensar en ellos como una serie de decisiones organizadas en una estructura de árbol, siendo los nodos del árbol preguntas sobre una característica específica de los datos, las ramas del árbol las posibles respuestas a esta y las hojas del árbol las decisiones o predicciones finales. Aunque, generalmente están sobre ajustados a los datos de entrenamiento y no se generalizan bien en pruebas independientes.

Tabla 3. Clasificación del pH de un suelo agrícola. Fuente: Villarroel 1988

Clasificación	pH
Muy fuertemente ácido	< 4,5
Fuertemente ácido	4,6 – 5,2
Moderadamente ácido	5,3 – 5,9
Débilmente ácido	6,0 – 6,5
Neutro	6,6 – 7,0
Débilmente alcalino	7,1 – 7,5
Moderadamente alcalino	7,6 – 8,0
Fuertemente alcalino	8,1 – 9,0
Muy fuertemente alcalino	> 9,0

Un enfoque potencialmente más rápido, pero menos preciso es evaluar los atributos individualmente y ordenarlos, descartando los atributos que caen por debajo de un punto de corte elegido. Por esto, se realizará una selección de atributos de las propiedades edáficas y de las longitudes de onda para saber cuál tienen mayor relevancia en la caracterización de las muestras del suelo, para posteriormente hacer un árbol de decisión con la clasificación filtrada y compararla con la clasificación sin filtrar y así saber que modelo se ajusta más a la realidad y cual viene mejor a nuestro estudio. Utilizando para ello como atributo evaluador J48 y *Random Forest*, este último aplicado únicamente a las longitudes de onda, con las técnicas *training set*, *cv 10 folds* y *Split* al 66%.

4. Resultados

En este apartado se presenta en primer lugar un análisis estadístico de las propiedades edáficas empleadas en este estudio. Seguidamente, se muestran los resultados de la modelización de las propiedades edáficas mediante radiometría proximal VNIR. Finalmente, se presentan los resultados de la aplicación de las técnicas de inteligencia artificial para la predicción de niveles de degradación de propiedades edáficas en función de espectros VNIR.

4.1. Análisis de las propiedades edáficas del área de estudio.

Una vez procesados los datos, obtenemos resultados de las propiedades edáficas de las poblaciones, tales como su materia orgánica, pH, conductividad eléctrica, nitrógeno, fósforo y potasio (Tabla 4). Previamente se aplicó el test de Shapiro-Wilk para analizar la distribución estadística de las variables. Los resultados de dicho test revelaron que las variables CE, NTK, P y K, no presentaban una distribución normal. Por ello, se decidió el adoptar test estadísticos no paramétricos en subsiguientes análisis. Se aplicó el test de Kruskal-Wallis para analizar posibles diferencias en los parámetros edáficos entre las diferentes poblaciones. Además, se realizaron diagramas de cajas (*box-plots*) para analizar gráficamente los valores paramétricos de las propiedades edáficas (Figura 8).

Tabla 4. Estadísticas descriptivas (mínimo, máximo, media y desviación estándar) de las propiedades edáficas para las diferentes poblaciones estudiadas. Se incluye el resultado (p-valor) del test de Kruskal-Wallis (K-W).

		Población										K-W (p-valor)
		1	2	3	4	5	6	7	8	9	10	
M.O. (%)	<i>Min</i>	1.88	2.80	2.95	2.41	1.77	2.36	2.56	0.38	2.91	0.56	<0,001 ***
	<i>Max</i>	6.06	11.84	7.69	8.78	4.10	6.77	9.50	6.67	8.45	2.67	
	<i>Media</i>	4.52	7.02	4.84	5.79	3.20	4.66	5.83	2.87	5.08	1.44	
	<i>DS</i>	1.34	2.96	1.79	1.85	0.80	1.52	2.19	2.45	1.55	0.66	
pH	<i>Min</i>	5.1	5.1	5.7	4.5	5.7	4.7	5.4	5.4	6.0	5.6	0,008 **
	<i>Max</i>	7.0	7.2	6.9	7.5	7.5	7.0	7.5	7.3	6.9	6.4	
	<i>Media</i>	5.7	6.3	6.2	6.6	6.5	6.0	6.3	6.5	6.5	6.1	
	<i>DS</i>	0.6	0.8	0.4	0.8	0.6	0.9	0.8	0.6	0.3	0.3	
CE (mS/cm)	<i>Min</i>	0.10	0.16	0.10	0.12	0.13	0.15	0.08	0.05	0.15	0.04	<0,001 ***
	<i>Max</i>	0.82	0.93	0.32	0.51	0.44	0.49	0.83	0.22	0.35	0.24	
	<i>Media</i>	0.29	0.45	0.16	0.37	0.33	0.33	0.35	0.15	0.21	0.10	
	<i>DS</i>	0.20	0.26	0.07	0.12	0.10	0.13	0.24	0.07	0.06	0.07	
NTK (%)	<i>Min</i>	0.08	0.09	0.07	0.03	0.05	0.08	0.09	0.03	0.08	0.02	<0,001 ***
	<i>Max</i>	0.21	0.41	0.23	0.32	0.13	0.21	0.27	0.18	0.25	0.09	
	<i>Media</i>	0.14	0.22	0.13	0.19	0.09	0.15	0.20	0.09	0.17	0.05	
	<i>DS</i>	0.04	0.12	0.05	0.09	0.02	0.04	0.07	0.06	0.04	0.02	
P (ppm)	<i>Min</i>	46.76	51.28	61.36	60.08	72.88	53.88	58.04	47.36	61.40	47.48	0,001 **
	<i>Max</i>	85.36	92.60	89.44	121.60	101.68	94.68	107.16	76.40	92.44	81.96	
	<i>Media</i>	63.18	70.37	73.34	89.16	85.90	70.79	73.68	66.70	69.84	68.27	
	<i>DS</i>	9.94	13.15	8.03	20.86	9.01	14.82	16.34	8.57	9.69	8.8	
K (ppm)	<i>Min</i>	281.9	279.9	322.0	453.6	260.5	197.3	428.6	90.9	1280.9	179.3	<0,001 ***
	<i>Max</i>	663.9	3711.2	737.1	2679.9	1181.1	2876.2	3948.7	847.5	2655.1	808.6	
	<i>Media</i>	503.3	1545.3	496.7	1628.9	710.8	998.8	2464.5	524.9	1799.4	368.8	
	<i>DS</i>	134.1	1290.0	140.8	648.6	269.4	975.99	1316.5	286.1	422.6	225.7	

Significatividad: [**] p-valor < 0,01 ; [***] p-valor < 0,001

En la Tabla 4 se muestran los estadísticos descriptivos de las propiedades edáficas y los resultados del test de Kruskal-Wallis. De ellas podemos destacar en la materia orgánica un máximo de 11,84% dentro del poblado 2 (Badiari) y un mínimo de 0,38% en el poblado 8 (Tanague). Un contenido de materia orgánica del 10 % en un suelo arenoso podrá generar una tonalidad bastante oscura, mientras que ese mismo contenido para en un suelo arcilloso apenas podrá generar un pequeño cambio en el color (García et al. 2020). Además, observamos la conductividad eléctrica de Badiari con 0,93 mS/cm y de Tanague con 0,05 mS/cm, indicándonos la diferencia en contenido de sales dado que su valor es menor a 0,001 indicándonos que tiene diferencias significativas.

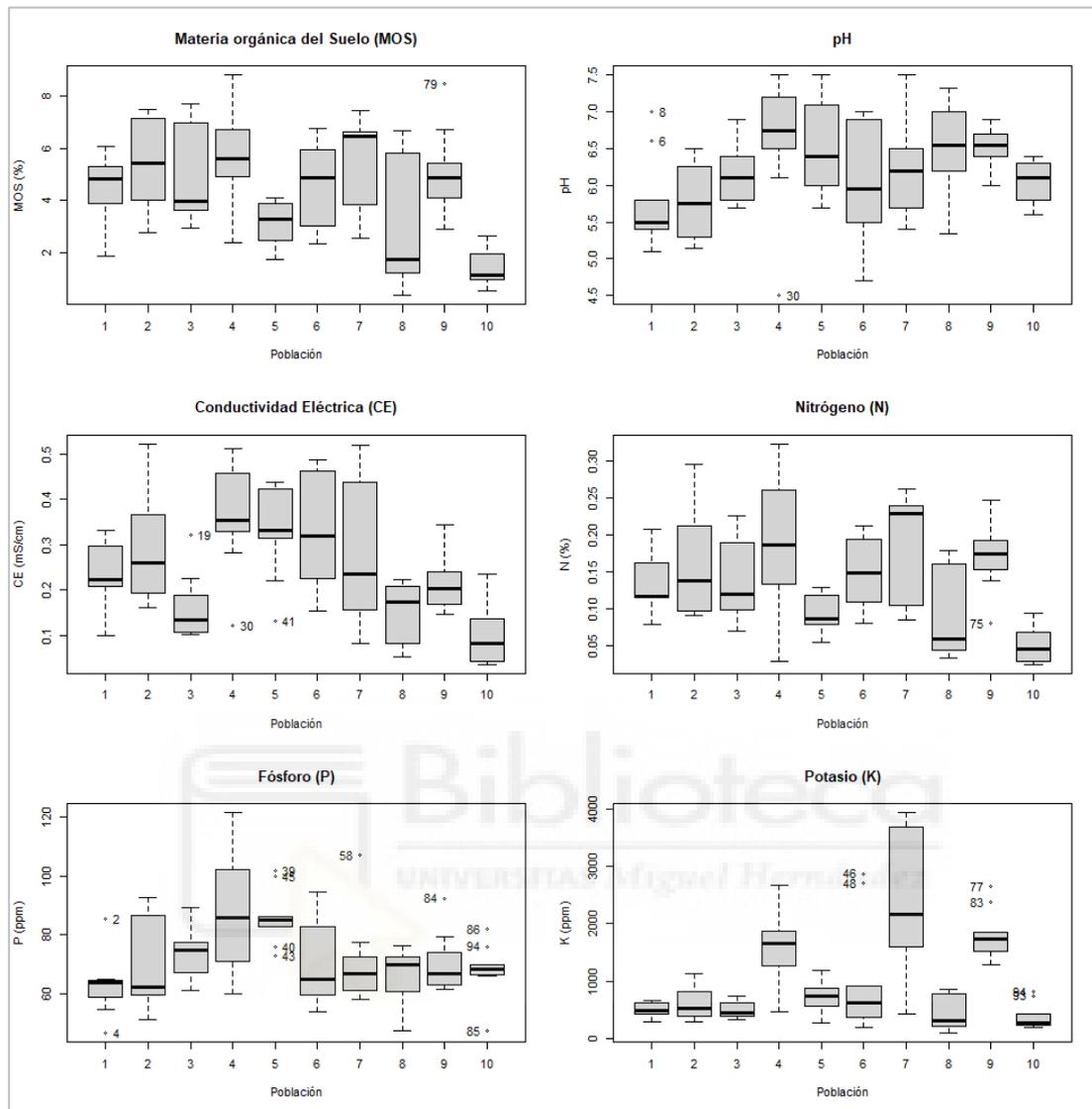


Figura 8. Diagramas de cajas (*box-plots*) de las propiedades edáficas para las diferentes poblaciones (números de población incluidos en Tabla 1).

Seguidamente se procedió a analizar la relación entre las diferentes variables edáficas, empleándose para ello el test de correlación no paramétrico de los rangos de Spearman (Tabla 5), destacando especialmente la relación entre las propiedades del contenido de nitrógeno, con el contenido de materia orgánica, la conductividad eléctrica y el potasio, las cuales presentan un aumento conjunto. En el caso de la materia orgánica, su relación sugiere una mayor disponibilidad de nutrientes y enriquecimiento del suelo. Por otro lado, en el caso de la conductividad eléctrica, esta relación podría reflejar una concentración superior nutrientes

disueltos en el suelo, incluido el nitrógeno, o bien podría ser un indicador de posibles problemas de salinidad si los niveles de CE son excesivamente altos.

Tabla 5. Resultados del test de correlación de Spearman. Se indican los niveles de significatividad de las correlaciones conforme al test de Holm.

	M.O.	pH	CE	NTK	P	K
M.O.	1	0.24	0.57***	0.90***	0.17	0.64***
pH	0.24	1	0.43***	0.22	0.40***	0.39***
CE	0.57***	0.43***	1	0.59***	0.37**	0.55***
NTK	0.90***	0.22	0.59***	1	0.16	0.68***
P	0.17	0.40***	0.37**	0.16	1	0.35**
K	0.64***	0.39***	0.55***	0.68***	0.35**	1

Significatividad: [**] p-valor < 0,01; [***] p-valor < 0,001

4.2. Determinación de propiedades edáficas mediante espectroscopía VNIR

La aplicación de los modelos PLSR para la estimación de propiedades edáficas usando técnicas de espectroscopia proximal es de gran relevancia para la evaluación de problemas de degradación de suelos (Dhawale et al. 2015). En este trabajo se realizaron modelos PLSR de calibración de nuestras muestras, donde las propiedades edáficas eran las variables dependientes y las longitudes de onda 480/860 las variables independientes (Tabla 6) mediante el método *kernelpls*.

Tabla 6. Modelo PLSR Calibración a longitudes onda 480/860 en Preloi 70/30. Fit method: kernelpls

Parámetros (Y)	Nº comp	Calibración		Validación					
		RMSE	R ²	RMSE	NRMSE	Mae	Cor	R ²	RPD
MOS	4	1856.0000	0.1985	1.7541	24.0590	1.4073	0.5235	0.2741	0.7397
pH	3	0.5811	0.1126	0.5583	19.9382	0.4631	0.6117	0.3742	0.7566
CE	3	0.1466	- 0.2448	0.1312	28.7459	0.1153	0.2705	0.0732	0.4720
NTK	4	0.0658	0.1062	0.0597	23.9697	0.0523	0.5642	0.3183	0.6326
P	3	14.7600	- 0.0977	16.5253	22.0808	13.0791	- 0.0355	0.0013	0.2082
K	2	887.2000	- 0.0176	868.8285	24.1227	623.6319	0.1369	0.0187	0.5256

Se utilizan 10 componentes y la validación es "*Leave One Out*" (LOO), con lo que evaluamos el rendimiento del PLSR en relación con las variables predictoras (longitudes de onda) y variables objetivo (propiedades edáficas). Como se mencionó anteriormente, la elección del número de componentes, y se escogió en función de, conseguir el menor error posible entre observados y predichos del coeficiente de variación (cv), el mayor porcentaje posible de varianza y el R^2 lo más próximo posible a 1. Por esta razón, se escogió el siguiente número de componentes para cada propiedad edáfica: materia orgánica (4), pH. (3), conductividad (3), nitrógeno (4), fósforo (3) y potasio (2). Podremos apreciar mejor la calidad de las predicciones del modelo en la figura 9 donde compara los valores observados frente a los predichos de cada variable, cuanto más se acerque nuestro modelo a la línea central, mejor habrá sido la precisión de este.



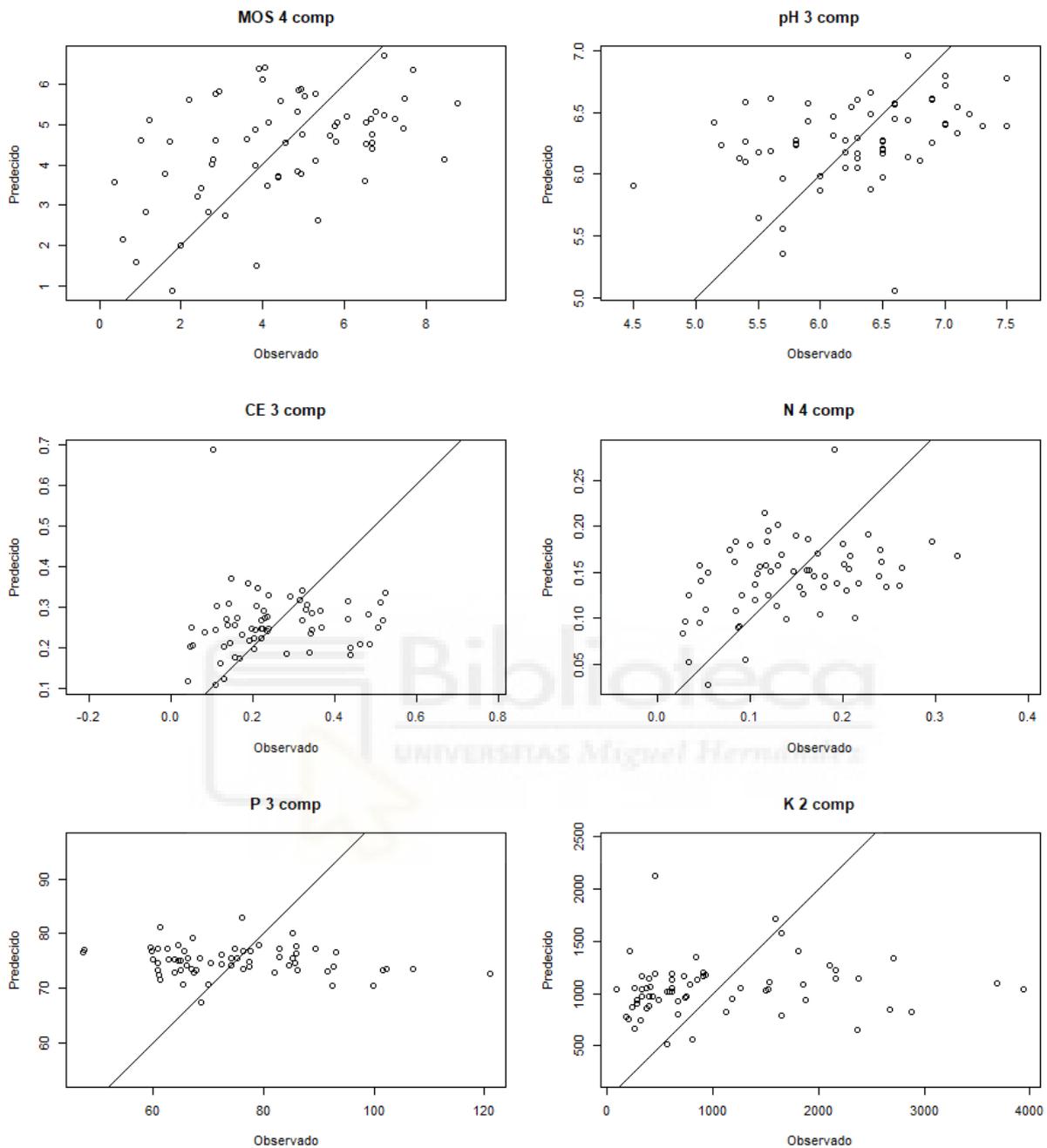


Figura 9. Diagramas de dispersión de los modelos PLSR de calibración para las propiedades edáficas.

Una vez calibrado nuestro modelo, procede a ser validado con el número de componentes que hemos elegido, con el que conseguimos nuevamente una tabla de descriptivos básicos de los valores observados frente a los predichos (Tabla 7) en función de los componentes. Posterior a esto, se realizan una serie de test tales como el error cuadrático medio (RMSE), el

error cuadrático medio normalizado (NRMSE), y Error Absoluto Medio (MAE) que muestra el error de los valores observados frente los predichos, además se calculara el Ratio de Rendimiento a Desviación (RPD) y un test de correlación y R^2 . Primeramente, observamos que los valores de RMSE y MAE son similares, por lo que comprobamos que los test han salido correctamente, además de que cuanto más bajo más se precisos son porque se acercan a los valores reales. Observamos el nivel de correlación es positivo por lo que existe una relación lineal positiva fuerte entre predicciones y valores reales. Además, comparamos R^2 medido en calibración y validación, que cuanto más cercano a 1 mejor se explican la variación de datos, siendo mayor el obtenido en validación. Por último, un RPD alto indica que las predicciones del modelo son muy precisas en comparación con la variación natural de los datos.

Tabla 7. Modelo PLSR Calibración a longitudes onda 480/860 en Preloi 70/30. Fit method: kernelp

	RMSE	NRMSE	Mae	Cor	R^2 val	RPD
MOS	1,754144	24,05903	1,407317	0,523502	0,274054	0,739696
pH	0,58406	20,85939	0,49591	0,56374	0,317801	0,64029
CE	0,13226	28,9794	0,11606	0,26104	0,068143	0,49489
NTK	0,059658	23,96971	0,052338	0,564156	0,318272	0,632643
P	16,1498	21,57913	12,9914	-0,1708	0,029159	0,11119
K	847,757	23,53769	655,722	0,06733	0,004534	0,32929

Generamos gráficos de dispersión comparando valores observados y valores predichos por el modelo PLSR, evaluando visualmente qué tan bien se ajustan las predicciones del modelo a los valores reales para diferentes variables. Los gráficos están separados por variable y muestran una línea de referencia para evaluar la relación entre los valores observados y los predichos, que cuanto más cercana estén a ella, mejor es la predicción.

Se realizó nuevamente gráficos de predicción (Figura 10) manipulando los datos de validación y el modelo PLSR, comparando los valores observados y valores predichos en el conjunto de validación permitiendo evaluar la calidad de las predicciones del modelo para diferentes variables, de nuevo con el objetivo de que se ajuste lo más posible a la línea de referencia.

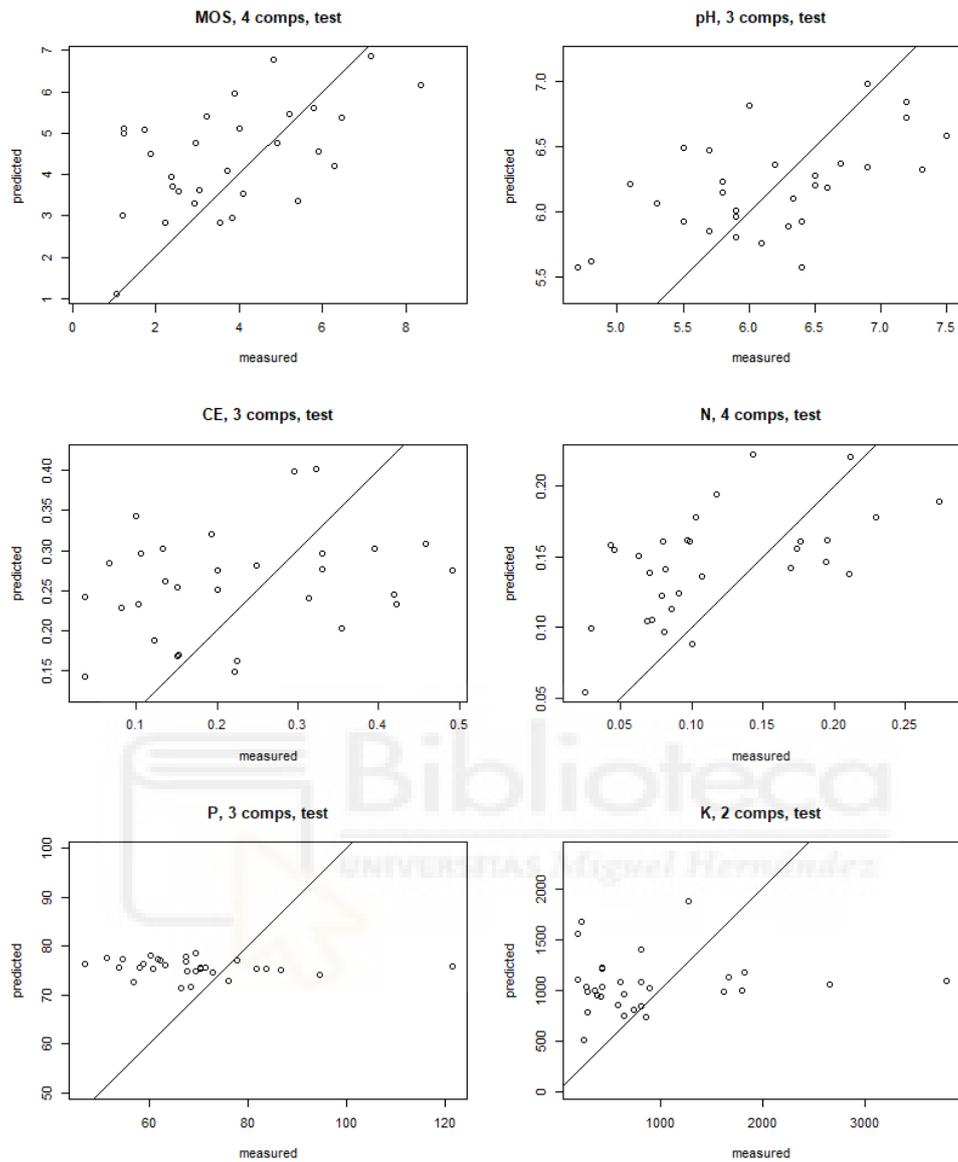


Figura 10. Diagramas de dispersión de los modelos PLSR de validación para las propiedades edáficas.

4.3. Aplicación conjunta de técnicas de *machine learning* y radiometría VNIR para evaluar el nivel de degradación de suelos.

Realizada las clasificaciones, observamos que, dependiendo de la propiedad utilizada para categorizar el suelo, ciertas longitudes de onda y propiedades destacan por tener mayor relevancia sobre las demás (Tabla 8).

Tabla 8. importancia longitudes de onda

	VNIR en orden importancia mayor a menor									
MOS_cat	vnir960	vnir583	vnir585	vnir586	vnir587	vnir588	vnir589	vnir584	vnir582	vnir591
pH_cat	vnir960	vnir583	vnir585	vnir586	vnir587	vnir588	vnir589	vnir584	vnir582	vnir591

Una vez realizada la selección de atributos, se procede a clasificar y comparar las distintas propiedades y longitudes de onda para ver cual se asemeja más a la realidad y poder tener así un buen modelo para predecir las distintas propiedades del suelo y reflectancia, obteniendo como resultado lo adjunto a la Tabla 9, donde se muestra que en las propiedades edáficas prácticamente no hay diferencia en haber filtrado o no las propiedades utilizadas y que todas las clasificaciones se asemejan bastante a la realidad, siendo las mejores a clasificaciones a utilizar la técnica del *training set* para la materia orgánica, y *split* al 66% para el pH.

Tabla 9. Clasificación de las propiedades edáficas

J48	correcto al... (%)		
	training set	cv 10 folds	split 66%
SOM/pH/CE/NPK			
MOS_cat	100	93,6	84,4
pH_cat	98,9	97,9	100
SOM/pH/CE			
MOS_cat	100	95,7	84,4
pH_cat	98,9	97,9	100

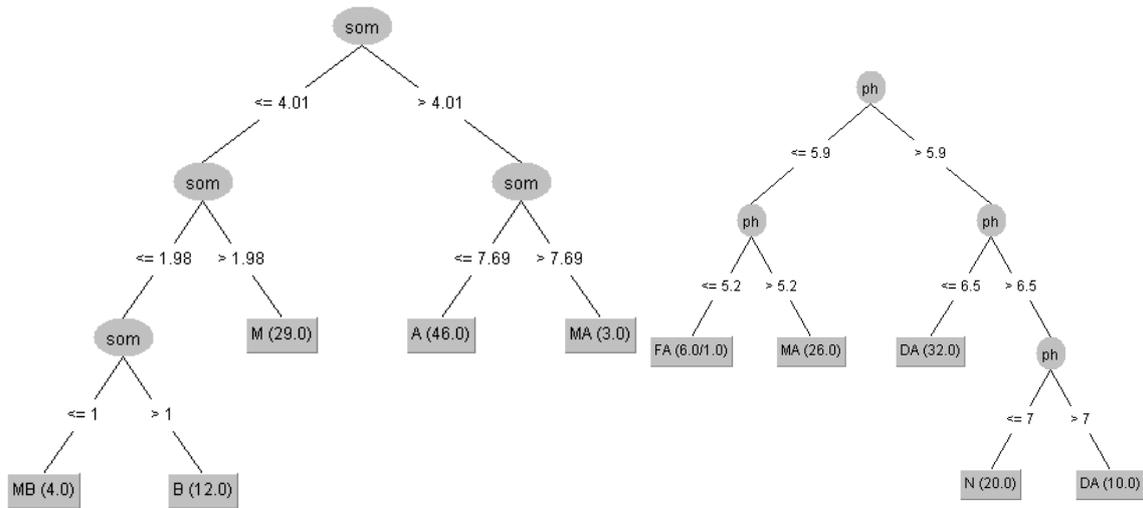


Figura X. Clasificación de las propiedades MOS (izquierda) con training set y pH (derecha) con Split 66%

En la tabla 10 adjunta, se observa cómo queda la clasificación de las longitudes de onda y cuanto se asemeja los modelos a los datos reales de manera bruta con todas las longitudes de onda, y filtrada usando solamente las 10 más relevantes. Se observa una clara diferencia en el atributo evaluador J48 cuando filtramos las longitudes de onda a cuando no, asemejándose más a la realidad cuando no hacemos un filtrado, pero en el atributo evaluador *Random Forest*, aunque es cierto que asemeja más a los valores reales cuando no se hace filtrado, no hay gran diferencia a cuando se filtra usando solo las más relevantes para hacer más sencillo el modelo, siendo el *training set* el que proyectó mejores resultados (Figura 11).

Tabla 10. Clasificación VNIR de las longitudes de onda

correcto al... (%)	J48			Random Forest		
	training set	cv 10 folds	split 66%	training set	cv 10 folds	split 66%
VNIR filtrado						
MOS_cat	75,5	51	46,8	100	43,6	50
pH_cat	61,7	39,4	37,5	100	30,6	37,5
VNIR sin filtrar						
MOS_cat	94,7	41,5	40	100	46,8	56,25
pH_cat	93,6	27,7	28,1	100	37,2	50

6. Bibliografía

Allory, V., Cambou, A., Moulin, P., Schwartz, C., Cannavo, P., Vidal-Beaudet, L. y Barthès, B.G. (2019). *Quantification of soil organic carbon stock in urban soils using visible and near infrared reflectance spectroscopy (VNIRS) in situ or in laboratory conditions. Science of The Total Environment*, 686, 764-773. <https://doi.org/10.1016/j.scitotenv.2019.05.192>

Alomar, D., y Fuchslocher, R. (1998). FUNDAMENTOS DE LA ESPECTROSCOPIA DE REFLECTANCIA EN EL INFRAROJO CERCANO (NIRS) COMO MÉTODO DE ANÁLISIS DE FORRAJES. *Agro Sur*, 26(1), 88–104. <https://doi.org/10.4206/agrosur.1998.v26n1-11>

Brown, D.J., Brickley, R.S. y Miller, P.R. (2005). *Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma*, 129 (3-4), 251-267. <https://doi.org/10.1016/j.geoderma.2005.01.001>

Brodský, L., Vašát, R., Klement, A., Zádorová, T. y Jakšík, O. (2013). *Uncertainty propagation in VNIR reflectance spectroscopy soil organic carbon mapping. Geoderma*, 199, 54-63. <https://doi.org/10.1016/j.geoderma.2012.11.006>

Brown, D.J. (2007). *Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. Geoderma*, 140(4), 444-453. <https://doi.org/10.1016/j.geoderma.2007.04.021>

Brunatti, C., y Martín, A. (2010). *Introducción a la espectroscopía de absorción molecular ultravioleta, visible e infrarrojo cercano*. Recuperado el, 24.

Burbano Orjuela, H. (2013). EL SUELO AL SERVICIO DE LA SOCIEDAD Y SU ROL EN EL CONTEXTO DE LOS CAMBIOS GLOBALES. *Tendencias*, 11(2), 53–62. Recuperado a partir de <https://revistas.udenar.edu.co/index.php/rtend/article/view/548>

Cambou, A., Allory, V., Cardinael, R., Carvalho Vieira, L. y Barthès, B.G. (2021). *Comparison of soil organic carbon stocks predicted using visible and near infrared reflectance (VNIR) spectra acquired in situ vs. on sieved dried samples: Synthesis of different studies. Soil Security*, 5. <https://doi.org/10.1016/j.soisec.2021.100024>

Cambou, A., Cardinael, R., Kouakoua, E., Villeneuve, M., Durand, C. y Barthès, B.G. (2016). *Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field. Geoderma*, 261, 151-159. <https://doi.org/10.1016/j.geoderma.2015.07.007>

Cortez, P. M. (2017). *ESPECTROSCOPIA DE INFRARROJO PARA TODOS... y 51 espectros de alimentos consumidos en México*. México. Recuperado de https://ciatej.mx/files/divulgacion/divulgacion_5a43b7c09fdc1.pdf.

Dhawale, N.M., Adamchuk, V.I., Prasher, S.O., Viscarra Rossel, R.A., Ismail, A.A., Kaur, J. (2015). *Proximal soil sensing of soil texture and organic matter with a prototype portable mid-infrared spectrometer*. *European Journal of Soil Science*, 66: 661-669

Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). *The WEKA Workbench*. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

Forcadell Pérez, G. J. (2022). *Instrumentación y medida de parámetros edáficos mediante el uso de sensores* [Trabajo Fin de Grado, Universidad Miguel Hernández]

Gallego Moreno, C. (2019). *Medio rural y gestión sostenible de suelos en la Reserva Natural Comunitaria de Dindéfelo, Senegal* [Trabajo Fin de Máster, Universidad Miguel Hernández]

García Orenes, M.F., Gómez Lucas, I., Jordán Vidal, M.M., Mataix Solera, J., Navarro Pedreño, J. (2020). *Edafología*

García Orenes, M.F., Gómez Lucas, I. (2022). *Degradación y Recuperación de Suelos*

Gee, G.W., Bauder, J.W., (1986). *Particle-size analysis*. In: Klute, A. (Ed.), *Methods of Soil Analysis. Part 1: Physical and Mineralogical Methods*. American Society of Agronomy (ASA) and Soil Science Society of America (SSSA), Madison (WI), USA, pp. 383–411

Gholizadeh, A., Saberioon, M., Pouladi, N. y Ben-Dor, E. (2023). *Quantification and depth distribution analysis of carbon to nitrogen ratio in forest soils using reflectance spectroscopy*. *International Soil and Water Conservation Research*, 11(1), 112-124. <https://doi.org/10.1016/j.iswcr.2022.06.004>

Gómez, R., y Murillo, R. (2006). *Espectroscopía infrarroja*. Universidad Nacional Autónoma de México.

González Corchón, C. (2023). *Ficha sector: Tecnología agrícola en Senegal 2023*. https://www.icex.es/content/dam/es/icex/oficinas/038/documentos/2023/04/ficha-sector/FS_Tecnolog%C3%ADa%20agr%C3%ADcola%20en%20Senegal%202023_REV.pdf

Graciani, S., Del Barco, P. y Stangaferro, J. (2011). PRINCIPIOS FÍSICOS DE TELEDETECCIÓN. <http://www.bibliotecacpa.org.ar/greenstone/collect/facagr/index/assoc/HASH3aa3.dir/doc.pdf>

Houba, V.J.G., Temminghoff, E.J.M., Gaikhorst, G.A., van Vark, W. (2000). *Soil analysis procedures using 0.01 M calcium chloride as extraction reagent. Communications in Soil Science and Plant Analysis*, 31: 1299-1396

ICEX-INVEST, I. S. (2015). España, plataforma para las inversiones y sedes de empresas multilatinas en Europa, África y Oriente Medio.

Jokoo Int. (s.f.). Plan Senegal Emergente. https://www.jokooint.com/images/senegal_files/PlanSenegalEmergente.pdf

Knox, N.M., Grunwald, S., McDowell, M.L., Bruland, G.L., Myers, D.B. y Harris, W.G. (2015). *Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. Geoderma*, 239-240, 229-239. <https://doi.org/10.1016/j.geoderma.2014.10.019>

Meléndez Pastor, I., Navarro Pedreño, J., Gómez, I. y Koch, M. (2008). *Identifying optimal spectral bands to assess soil properties with VNIR radiometry in semi-arid soils. Geoderma*, 147(3-4), 126-132. <https://doi.org/10.1016/j.geoderma.2008.08.004>

Mevik, B.-H.; Wehrens, R. (2007). The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, 18: 1–23.

Olatunde, K.A. (2021). *Determination of petroleum hydrocarbon contamination in soil using VNIR, DRS and PLSR modeling. Heliyon*, 7(4). <https://doi.org/10.1016/j.heliyon.2021.e06794>

Qinghu Jiang, Qianxi Li, XinggangWanga, YuWua, Xiaolu Yang, Feng Liu. (2017). *Estimation of soil organic carbon and total nitrogen in different soil layers. Geoderma*, 293, 54-63. <https://doi.org/10.1016/j.geoderma.2017.01.030>

Rojas, A. E., y Ibarra, J. (2003). *La degradación del suelo y sus efectos sobre la población. Población y desarrollo*, (25), 5-10.

The Global Risks Report 2023. (2023). World Economic Forum. <https://www.weforum.org/reports/global-risks-report-2023/>

Tsakiridis, N.L., Theocharis, J.B., Ben-Dor, E. y Zalidis, G.C. (2019). *Using interpretable fuzzy rule-based models for the estimation of soil organic carbon from VNIR/SWIR spectra and soil*

texture. *Chemometrics and Intelligent Laboratory Systems*, 189, 39-55.
<https://doi.org/10.1016/j.chemolab.2019.03.011>

Tsakiridis, N.L., Theocharis, J.B., Symeonidis, A.L. y Zalidis, G.C. (2021). *Improving the predictions of soil properties from VNIR–SWIR spectra in an unlabeled region using semi-supervised and active learning*. *Geoderma*, 387.
<https://doi.org/10.1016/j.geoderma.2020.114830>

Valenciaga, D., y Oliveira Simoes Saliba, E. D. (2006). *La espectroscopia de reflectancia en el infrarrojo cercano (NIRS) y sus potencialidades para la evaluación de forrajes*. *Revista Cubana de Ciencia Agrícola*, 40(3), 259-267.

Valeriano Martinez Garcia. (2019). Comparación de dos aproximaciones para la medida del carbono orgánico del suelo: un método de referencia en el laboratorio y el método por infrarrojo cercano (NIR) [Trabajo Fin de Grado, Universidad Miguel Hernández]

Wei Tan, Xiaorong Qin, Jiacheng Liu, Joseph Michalski, Hongping He, Yuzeng Yao, Meijun Yang, Jian Huang, Xiaojun Lin, Chaoqun Zhang y Xiaoliang Liang. (2021). *Visible/near infrared reflectance (VNIR) spectral features of ion-exchangeable Rare earth elements hosted by clay minerals: Potential use for exploration of regolith-hosted REE deposits*. *Applied Clay Science*, 215. <https://doi.org/10.1016/j.clay.2021.106320>

Wei Tan, Xiaoyan Wang, Hongyan He, Wenwen Qi. (2022). Absolute radiometric calibration with lunar surface reflectance as reference for earth-observing VNIR camera. *Heliyon*, 8(12).
<https://doi.org/10.1016/j.heliyon.2022.e11776>

Wold, S.; Sjostrom, M.; Eriksson, L. (2001) PLS-regression: A basic tool of chemometrics. *Chemometrics Intelligent Laboratory Systems*, 58: 109–130.

World Bank. (2008). *Senegal: Country Environmental Analysis*. © Washington, DC. <http://hdl.handle.net/10986/7773> License: CC BY 3.0 IGO.

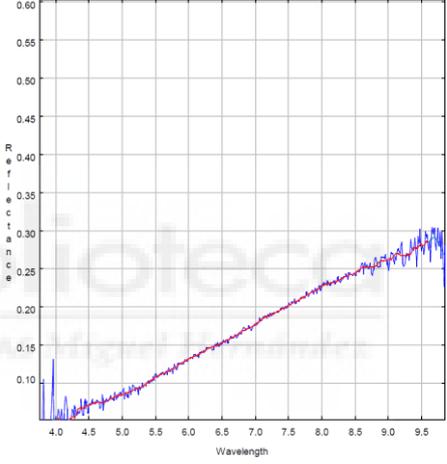
Xiang Wang, Lin Li, Huanjun Liu, Kaishan Song, Liping Wang y Xiangtian Meng. (2022). *Prediction of soil organic matter using VNIR spectral parameters extracted from shape characteristics*. *Soil and Tillage Research*, 216. <https://doi.org/10.1016/j.still.2021.105241>

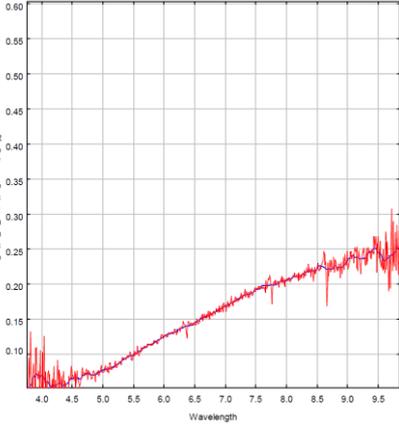
Fuentes web:

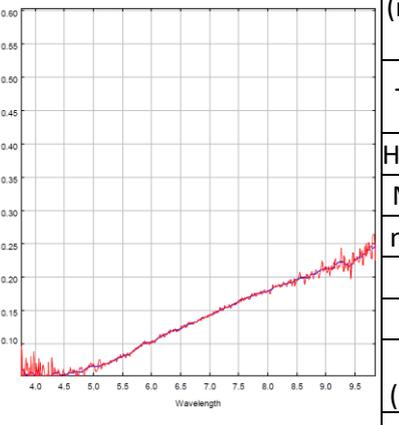
- Adagio Consultoría Tecnológica S.L. (2023). *Senegal: Geografía*. <https://senegambia.es/sen-geografia>
- Colonel Baba BA, ~~Head of Reforesting Department, Senegal~~. (s.f.). *Land Degradation and Good Practices in Senegal*. <http://aicd-africa.org/archives/818>
- GeoNames. (s.f.) <https://www.geonames.org/>
- IBM. (s.f.). *What is machine learning?* <https://www.ibm.com/topics/machine-learning>
- Organización de Naciones Unidas para la Alimentación y Agricultura. (29 de noviembre de 2019). *Evaluación de la Degradación de Tierras a Nivel Local*. <https://www.fao.org/soils-portal/soil-degradation-restoration/evaluacion-a-nivel-local/es/>
- RDT Simulation. (s.f.). *CLASIFICACIÓN CLIMÁTICA MUNDIAL SEGÚN KÖPPEN GEIGER. Ingeniería para Arquitectura Bioclimática y Sostenible*. <https://www.simulacionesyproyectos.com/blog-ingenieria-arquitectura/clasificacion-climatica/>
- World Bank Group. (s.f.). *Climate Change Knowledge Portal for Development Practitioners and Policy Makers*. <https://climateknowledgeportal.worldbank.org/country/senegal/climate-data-historical>

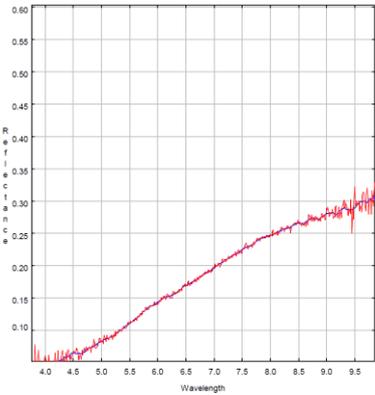
7. Anexos

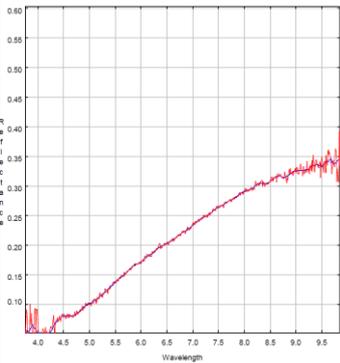
7.1. Librería de suelos

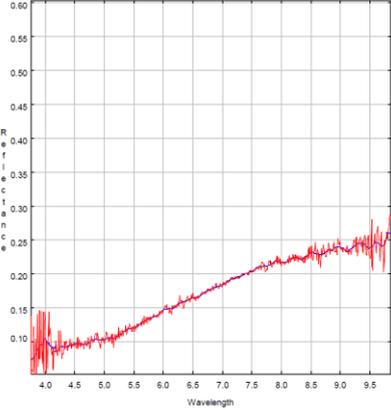
	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	1-Afia
				Muestra	10
Circulo escala				Color (muestra seca)	10YR 4/2
				Textura	Areno-franca
				Humedad	0.99
				MOS(%)	4.89
				mos_cat	A
				pH	5.8
				ph_cat	MA
				CE (mS/cm)	0.24
				ce_cat	NS
				N (%)	0.16
			P (ppm)	65.08	
			K (ppm)	488.47	

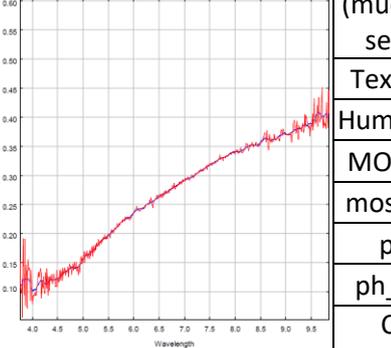
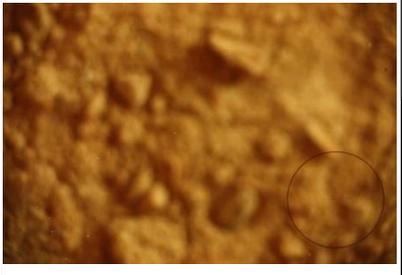
	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	2-Badiari
				Muestra	18
Círculo escala				Color (muestra seca)	7,5YR 4/3
				Textura	Arenosa
				Humedad	0.71
				MOS(%)	4.01
				mos_cat	M
				pH	5.5
				ph_cat	MA
				CE (mS/cm)	0.19
				ce_cat	NS
				N (%)	0.10
			P (ppm)	61.80	
			K (ppm)	383.42	

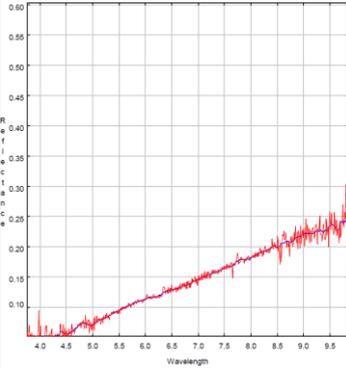
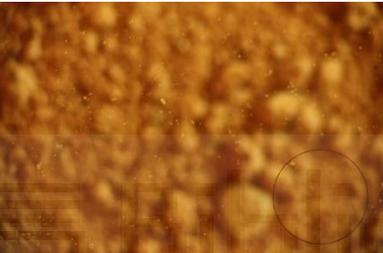
	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	3-Dande
				Muestra	25
Círculo escala				Color (muestra seca)	7,5YR 3/3
				Textura	Franco-arenosa
				Humedad	0.56
				MOS(%)	3.81
				mos_cat	M
				pH	6.9
				ph_cat	N
				CE (mS/cm)	0.23
				ce_cat	NS
				N (%)	0.12
			P (ppm)	89.44	
			K (ppm)	393.38	

	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	4- Goumbambere
				Muestra	39
Circulo escala				Color (muestra seca)	5YR 3/4
				Textura	Areno-franca
				Humedad	1.63
				MOS(%)	8.78
				mos_cat	MA
				pH	6.5
				ph_cat	DA
				CE (mS/cm)	0.51
				ce_cat	NS
				N (%)	0.32
			P (ppm)	60.08	
			K (ppm)	2366.83	

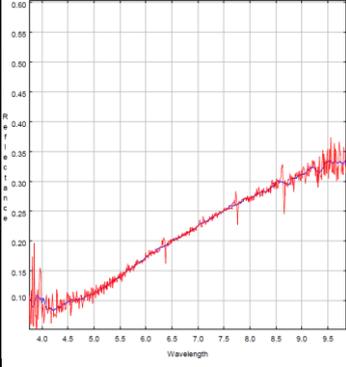
	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	5- Koukoudji
				Muestra	44
Circulo escala				Color (muestra seca)	10YR 4/3
				Textura	Franco-arenosa
				Humedad	0.57
				MOS(%)	2.84
				mos_cat	M
				pH	5.7
				ph_cat	MA
				CE (mS/cm)	0.32
				ce_cat	NS
				N (%)	0.08
			P (ppm)	101.68	
			K (ppm)	365.94	

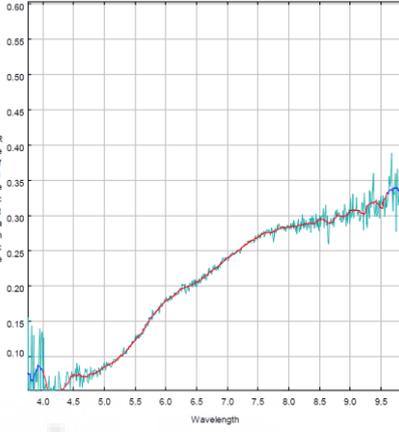
	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	6-Nandoumari
				Muestra	53
Circulo escala				Color (muestra seca)	7,5YR 4/2
				Textura	Arenosa
				Humedad	1.11
				MOS(%)	2.84
				mos_cat	M
				pH	5.9
				ph_cat	MA
				CE (mS/cm)	0.16
				ce_cat	NS
				N (%)	0.11
			P (ppm)	82.88	
			K (ppm)	2709.88	

	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	7-Nepen Peuhl
				Muestra	69
Circulo escala				Color (muestra seca)	10YR 5/4
				Textura	Arenosa
				Humedad	0.52
				MOS(%)	2.75
				mos_cat	M
				pH	5.5
				ph_cat	MA
				CE (mS/cm)	0.16
				ce_cat	NS
				N (%)	0.11
			P (ppm)	65.48	
			K (ppm)	1592.59	

	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	8-Tanague
				Muestra	74
Circulo escala				Color (muestra seca)	10YR 4/4
				Textura	Franco-arenosa
				Humedad	0.57
				MOS(%)	2.18
				mos_cat	M
				pH	7.1
				ph_cat	DA
				CE (mS/cm)	0.21
				ce_cat	NS
				N (%)	0.08
			P (ppm)	74.12	
			K (ppm)	393.47	

UNIVERSITAS Miguel Hernández

	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	9-Thiangue
				Muestra	89
Circulo escala				Color (muestra seca)	7,5YR 3/3
				Textura	Franco-arenosa
				Humedad	0.85
				MOS(%)	4.86
				mos_cat	A
				pH	6.5
				ph_cat	DA
				CE (mS/cm)	0.24
				ce_cat	NS
				N (%)	0.19
			P (ppm)	74.24	
			K (ppm)	2379.81	

	Suelo seco	Suelo tras LOI	Espectro		
Enfoque normal				Población	10-Teppere
				Muestra	100
Circulo escala				Color (muestra seca)	5YR 4/4
				Textura	Arenosa
				Humedad	0.85
				MOS(%)	2.669
				mos_cat	M
				pH	6.1
				ph_cat	DA
				CE (mS/cm)	0.24
				ce_cat	NS
				N (%)	0.09
			P (ppm)	76.16	
			K (ppm)	808.64	

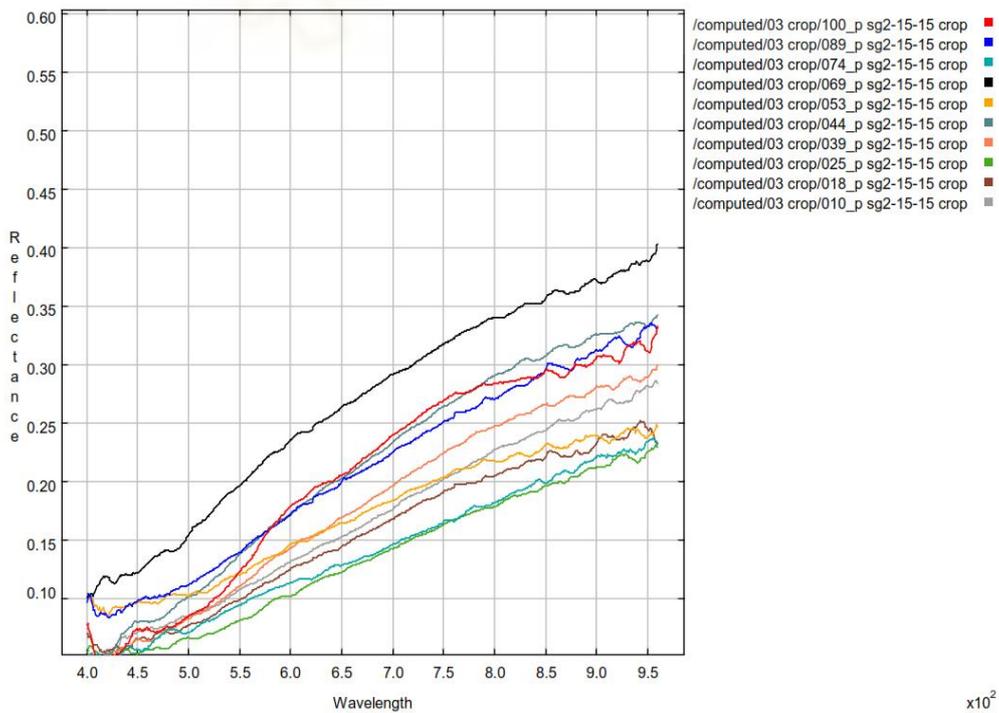


Figura 12. Comparativa espectros suelos