

UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE CIENCIAS



TESIS

**ANÁLISIS POR IMÁGENES HIPERESPECTRALES DE
MEZCLAS DE SUELO CON DERIVADOS DE PETRÓLEO,
EN LA REGIÓN DE 900 A 1700 nm**

PARA OBTENER EL GRADO ACADÉMICO DE MAESTRO EN CIENCIAS EN FÍSICA
ELABORADA POR:

RICARDO FLORES CAMARGO

ASESOR:

DR. HÉCTOR RAÚL LORO RAMÍREZ

LIMA-PERÚ

2019

Este trabajo le dedico a mis padres
Zulema Camargo Guerra
Sigifredo Flores Zapata

AGRADECIMIENTOS

Agradezco a todas las personas que estuvieron presente y me apoyaron en todo momento para iniciar y culminar la presente tesis, así mismo agradezco a mi asesor y al grupo de imágenes hiperspectrales, como a mis compañeros que compartimos la maestría. Al Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC), Fondo Nacional de Desarrollo Científico de Innovación Tecnológica (FONDECYT), convenio Nro 167 FONDECYT-UNI por la beca otorgada para estudios de maestría.

Índice

Abstract	14
INTRODUCCIÓN	16
Planteamiento del problema	18
Hipótesis	19
Justificación	19
Objetivos	20
Metodología	20
CAPÍTULO 1	22
1 Imágenes hiperespectrales, adquisición y procesamiento	22
1.1 Imágenes hiperespectrales (HSI)	22
1.2 Hipercubo	23
1.3 Tipos de sistemas de adquisición	24
1.3.1 Sistema de imagen hiperespectral de laboratorio	24
1.3.2 Sistema al aire libre	25
1.3.3 Sistema aerotransportado	25
1.4 Modo de escaneo y de detección	26
1.4.1 Modo de escaneo de una cámara hiperespectral	26
1.4.2 Modo de detección de una cámara hiperespectral	28
1.4.3 Componentes de una cámara hiperespectral	29
1.5 Calibración	30
1.6 Preprocesamiento espacial	31
1.6.1 Segmentación de imagen	31
1.6.2 Región de interés (ROIs)	32
1.6.3 Mosaicos	32
1.7 Preprocesamiento espectral	34
1.7.1 Normalización	35
1.7.2 Filtro suavizante Savitzky-Golay	36
1.7.3 Derivadas en el espectro de reflectancia	37
1.7.4 Corrección de línea base	37
1.7.5 Corrección de escala multiplicativa	38
1.8 Procesamiento	39
1.8.1 Análisis de componentes principales (PCA)	39
1.8.2 Prueba de hipótesis para clasificar imágenes hiperespectrales . .	41
1.9 Conceptos generales de espectrometría en el infrarrojo	42
1.9.1 Medidas en infrarrojo	44
CAPÍTULO 2	45
2 Suelos e hidrocarburos de petróleo	45
2.1 Suelos y propiedades espectrales del suelo	45
2.1.1 Componentes básicos del suelo	45
2.1.2 Propiedades físicas o mecánicas del suelo	45
2.1.3 Propiedades espectrales de los suelos	47

2.2	Hidrocarburos de petróleo (PHC)	48
2.2.1	Origen y propiedades	48
2.2.2	Contaminación del suelo por hidrocarburos de petróleo	49
2.2.3	Propiedades espectrales de los hidrocarburos de petróleo	49
CAPÍTULO 3		52
3	PROCEDIMIENTOS Y RESULTADOS	52
3.1	Calibración del sistema de imágenes hiperespectrales (HSI)	52
3.1.1	Reflectancia del teflón	52
3.1.2	Sistema de imagen hiperespectral de laboratorio	54
3.1.3	Calibración del sistema hiperespectral	56
3.2	Preparación de las muestras patrones de referencia (PR)	60
3.2.1	Muestras de suelo y derivados de petróleo puros	61
3.2.2	Patrones de referencia (PR)	61
3.3	Adquisición de imágenes	63
3.4	Preprocesamiento	64
3.4.1	Normalización de la matriz del hipercubo	65
3.4.2	Suavizado Savitzky-Golay, primera y segunda derivada	66
3.4.3	SNV y de-trend	68
3.4.4	Eliminación de fondo con PCA	71
3.4.5	Identificación de la máscara de trabajo	72
3.5	Procesamiento	73
3.5.1	Construcción de librería espectral	73
3.5.2	Análisis de las imágenes de suelos contaminados	78
3.5.3	Análisis del desplazamiento de los histogramas del componente principal (PC1)	80
3.5.4	Análisis de la banda de absorción a 1200 nm en función de la concentración del contaminante	91
CONCLUSIONES		105
BIBLIOGRAFIA		107
Anexo A Manejo del software SpectrononPro		112
Anexo B Códigos R		117

SÍMBOLOS Y ABREVIATURAS

- BIP: (Banda intercalada por píxel) del inglés (Band Interleaved by Pixel)
- BIL: (Banda intercalada por línea) del inglés (Band Interleaved by Line)
- BSQ: (Banda secuencial) del inglés (Band Sequential)
- CCD: (Dispositivo de carga acoplada) del inglés (Charge Coupled Device)
- CMOS: (Semiconductores de óxido metálico complementario) del inglés (Complementary Metal Oxide Semiconductor)
- DP: Derivado de Petróleo
- FOV: (Campo de visión) del inglés (Field Of View)
- FPS: (Fotogramas por segundos) del inglés (Frames Per Second)
- HSI: (Imágenes Hiperespectrales) del inglés (Hyperspectral Image)
- IFOV: (Campo de visión instantáneo) del inglés (Instantaneous Field Of View)
- Q1: Primer cuartil
- Q3: Tercer cuartil
- NA: (No disponible) del inglés (Not Available)
- NIR: (Infrarrojo cercano) del inglés (Nerard Infrared)
- PC: (Componente principal) del inglés (Principal Component)
- PCA: (Análisis de componentes principales) del inglés (Principal Component Analysis)
- PHC: (Hidrocarburo de Petróleo) del inglés (Petroleum Hydrocarbon)
- Píxel: (Elemento de imagen) del inglés (Picture Element)
- PR: Patrón de Referencia
- RI: Rango Intercuartílico
- ROI: (Región de interés) del inglés (Region Of Interest)
- SNV: (Variable normal estándar) del inglés (Standard Normal Variate)
- SNV-D: (Variable Normal Estándar y tendencia) del inglés (Standard Normal Variate and De-trending)
- λ : Longitud de onda
- %pp: Porcentaje de peso
- UAV: (Vehículo aéreo no tripulado) del inglés (Unmanned Aerial Vehicle)
- NDVI (Índice de vegetación de diferencia normalizada) del inglés (Normalized Difference Vegetation Index):

Índice de figuras

Figura 1.1	Imagen hiperespectral (Wang and Peng, 2018).	22
Figura 1.2	Hipercubo de una imagen multispectral, imagen hiperespectral y la matriz aleatoria. Adaptado de Falco (2015); Bajorski (2011).	23
Figura 1.3	Sistema de imagen hiperespectral de laboratorio donde se nombra las partes que lo caracterizan (cámara hiperespectral, lente, plataforma, muestra, motor, luces y computadora). Adaptado de Mishra et al. (2017).	24
Figura 1.4	Sistema al aire libre con fuente de luz natural. Adaptado de Mishra et al. (2017).	25
Figura 1.5	Sistema aerotransportado activo y pasivo (Analytik, 2019).	26
Figura 1.6	Cubo de datos que muestra la dimensión espacial (x,y) y la dimensión espectral (λ). Las regiones, etiquetadas A, B y C, ilustran la porción del cubo de datos que se adquiere durante un ciclo de adquisición (Li et al., 2018a).	27
Figura 1.7	Escáner de línea de barrido. Adaptado de Manolakis et al. (2016).	28
Figura 1.8	Modos de detección de una cámara hiperespectral. Adaptado de Castro (2015).	28
Figura 1.9	Componentes de una cámara hiperespectral para conformar un hipercubo. Adaptado de Manolakis et al. (2016).	30
Figura 1.10	Umbralización por medio de un histograma y enmascaramiento binario (Bajorski, 2011).	32
Figura 1.11	Mosaico de una imagen hiperespectral de cerezas (Li et al., 2018b).	33
Figura 1.12	Metodología en el análisis de imágenes hiperespectrales (HSI) para suelos contaminados con derivado de petróleo (DP).	34
Figura 1.13	Corrección Savitzky para un espectro de NDVI.	36
Figura 1.14	a) Espectro de reflectancia, b) derivada del espectro de reflectancia.	37
Figura 1.15	Cambio de línea base por de-trend (MathWorks, 2019).	38
Figura 1.16	PCA como una proyección ortogonal en una dimensión inferior: puntos proyectados en una línea (Manolakis et al., 2003).	41
Figura 1.17	Prueba hipótesis entre suelo limoso y el suelo limoso con crudo de petróleo a 12%pp. Se muestran las funciones de densidad y la intersección entre ellas.	42
Figura 1.18	Espectro electromagnético. Adaptado de Hecht (2000).	43
Figura 2.1	triángulo de textura de suelo diseñado por USDA (Gisbert Blanquer et al., 2010).	46
Figura 2.2	Reflectancia típica de cinco tipos de suelo: (E) orgánico, (B) orgánico mínimamente alterado, (C) afectado por hierro, (A) orgánico afectado y (D) con dominio de hierro. Adaptado de Winkelmann (2005).	48
Figura 2.3	La radiación infrarroja absorbida por una molécula causa que los enlaces individuales vibren en gran aproximación como osciladores diatómicos	50
Figura 2.4	Bandas de absorción en el NIR (Martínez, 2015).	50
Figura 2.5	Características de un espectro en el infrarrojo cercano (NIR) de un suelo contaminado con hidrocarburo. Adaptado de Winkelmann (2005).	51
Figura 3.1	Equipos utilizados para conocer la intensidad de luz reflejada por el teflón.	52
Figura 3.2	A. Esquema para medir la intensidad de luz incidente, B. Esquema para medir la intensidad de luz reflejada por el teflón.	53

Figura 3.3	Sistema hiperespectral con cámara Pika NIR-320 de la marca RESO-NON.	54
Figura 3.4	a) Patrón de teflón para $E_{referencia}$ b) Cubierta de lente para $E_{corriente-oscura}$	56
Figura 3.5	Patrón de referencia.	57
Figura 3.6	Espectro de teflón (reflectancia en unidades arbitrarias) con fuente luminosa de lámpara de tungsteno , monitor de PC y luz fluorescente.	58
Figura 3.7	Espectro de teflón (reflectancia en unidades arbitrarias) en un ambiente aislado de luces externas y con fuente luminosa halógena.	58
Figura 3.8	Espectro de teflón en un ambiente aislado y con fuente luminosa halógena con 5 minutos de diferencia entre un espectro y otro.	59
Figura 3.9	(PARTE SUPERIOR) Diagrama de cajas de datos de reflectancia de las 168 bandas con variación de 5.68 %, (PARTE INFERIOR) Diagrama de cajas de datos de reflectancia de las 136 bandas con variación de 1.69 %.	59
Figura 3.10	Esquema de la preparación de los patrones de referencia (PR) para los suelos mezclados con derivados de petróleo (DP).	60
Figura 3.11	Suelos y derivados de petróleo puros utilizados para tener los patrones de referencia.	61
Figura 3.12	(IZQUIERDA) PR de 70 % de arcilla y 30 %pp-crudo de petróleo, (DERECHA) Patrón de referencia de 70 % de arcilla y 30 %pp-diésel.	62
Figura 3.13	Patrones de referencia (PR) con crudo de petróleo, (SUPERIOR) suelos arenosos, (MEDIO) suelos limosos y (INFERIOR) suelos arcillosos.	62
Figura 3.14	Patrones de referencia (PR) con diésel, (SUPERIOR) suelos arenosos, (MEDIO) suelos limosos y (INFERIOR) suelos arcillosos.	63
Figura 3.15	Mosaico de los patrones de referencia de los suelos limosos con crudo de petróleo y los suelos limosos sin contaminar. La fila superior corresponde al mismo suelo utilizado como referencia en cada adquisición.	63
Figura 3.16	Espectros de los patrones de referencia de los suelo limosos con crudo de petróleo y los suelos limosos sin contaminar.	64
Figura 3.17	Partes del preprocesamiento.	65
Figura 3.18	(IZQUIERDA) Imagen normalizada en la banda 45 del suelo arcilloso, (DERECHA) Espectros de las 100 primeros píxeles que conforman la imagen hiperespectral de 40 000 píxeles.	65
Figura 3.19	Savitzky-Golay: (SUPERIOR) polinomio de grado 5 con 11 puntos, (INFERIOR-IZQUIERDA) polinomio de grado 5 con 25 puntos, (INFERIOR-DERECHA) polinomio de grado 5 con 47 puntos.	66
Figura 3.20	Primera derivada: (SUPERIOR) usando 11 puntos se observa una señal entre 1400 a 1500 nm, (INFERIOR-IZQUIERDA) usando 25 puntos se observa una mayor respuesta en la señal de 1400 a 1500 nm, (INFERIOR-DERECHA) con 47 puntos se observa una respuesta a la señal entre 1300 a 1400 nm.	67
Figura 3.21	Segunda derivada: (SUPERIOR) usando 11 puntos se observa que se pierde la señal, (INFERIOR-IZQUIERDA) usando 25 puntos se observa una señal de 1300 a 1500 nm, (INFERIOR-DERECHA) con 47 puntos se observa un pico máximo en 1450 nm.	68

Figura 3.22 Espectros de los 100 primeros píxeles del suelo arcilloso a los cuales se aplicó Nor-max/Sv-G/SNV.	69
Figura 3.23 Espectros de los 100 primeros píxeles del suelo arcilloso a los cuales se aplicó Nor-max/Sv-G/SNV-D.	69
Figura 3.24 Imagen en escala de grises de la banda 30 del suelo arcilloso sin contaminar a la cual se aplicó Nor-max/Sv-G/SNV.	70
Figura 3.25 Imagen en escala de grises de la banda 30 del suelo arcilloso sin contaminar a la cual se aplicó Nor-max/Sv-G/SNV-D	70
Figura 3.26 (IZQUIERDA) Scores de la primera componente de la muestra de suelo arcilloso, (DERECHA) histograma de los scores de la primera componente (N-max/Sv-G(5,11)/SNV/PC1).	71
Figura 3.27 Parte superior de derecha a izquierda hacia abajo: Imagen umbralizada con, histograma de la imagen, imagen de los píxeles de valores entre -4.52 y -2.54 Scores del PC1 y diagrama de caja.	72
Figura 3.28 Partes del procesamiento.	73
Figura 3.29 Espectros de reflectancia medios de las muestras de suelo limoso con diferentes concentraciones de crudo de petróleo.	74
Figura 3.30 Espectros de reflectancia medios de las muestras de suelo limoso con diferentes concentraciones de diésel.	75
Figura 3.31 Espectros de reflectancia medios de las muestras de suelo arcilloso con diferentes concentraciones de crudo de petróleo.	76
Figura 3.32 Espectros de reflectancia medios de las muestras de suelo arcilloso con diferentes concentraciones de diésel.	76
Figura 3.33 Espectros de reflectancia medios de las muestras de suelo arenoso con diferentes concentraciones de crudo de petróleo.	77
Figura 3.34 Espectros de reflectancia medios de las muestras de suelo arenoso con diferentes concentraciones de diésel.	77
Figura 3.35 Imagen de las primeras componentes de los seis (6) mosaicos que muestran los dos tipos de suelos contaminado con diferentes derivados de petróleo a diferentes niveles.	79
Figura 3.36 Funciones de distribución normal de los suelos limosos con crudo de petróleo a diferentes niveles de concentración frente al suelo limoso (blanco).	81
Figura 3.37 Funciones de distribución normal de los suelos arcillosos con crudo de petróleo a diferentes niveles de concentración frente al suelo arcilloso (blanco).	82
Figura 3.38 Funciones de distribución normal de los suelos arenosos con crudo de petróleo a diferentes niveles de concentración frente al suelo arenoso (blanco).	82
Figura 3.39 Función de densidad normal de los suelos limosos, arcillosos y arenosos contaminados con diésel frente al suelo no contaminado (blanco).	84
Figura 3.40 Función de densidad normal de los suelos limosos, arcillosos y arenosos contaminados con diésel frente al suelo no contaminado (blanco).	84
Figura 3.41 Función de densidad normal de los suelos limosos, arcillosos y arenosos contaminados con diésel frente al suelo no contaminado (blanco).	85
Figura 3.42 Distancia de los valores medios del blanco o muestra no contaminada a las muestras contaminadas o patrones de referencia.	88
Figura 3.43 Valores medios de las muestras contaminadas o patrones de referencia con su desviación estándar (sd).	89

Figura 3.44 Probabilidad de cometer el error tipo 1 cuando las concentraciones del suelo con contaminante cambian.	90
Figura 3.45 Clasificación de las muestras artificiales de suelo limoso con crudo de petróleo con 11 niveles de concentración por medio de los scores de la primera componente.	91
Figura 3.46 (IZQUIERDA) Espectros de reflectancia de los derivados del petróleo, (DERECHA) espectros de suelos.	92
Figura 3.47 Espectro de reflectancia SNV-D del suelo limoso a diferentes niveles de concentración con crudo de petróleo entre la región 1087 a 1374 nm. . . .	93
Figura 3.48 Segunda derivada de SNV-D del suelo limoso a diferentes niveles de concentración con crudo de petróleo entre la región 1160 a 1249 nm.	94
Figura 3.49 Regresión lineal entre las concentraciones y la segunda derivada de la reflectancia SNV-D del suelo limoso con crudo de petróleo en la longitud de onda 1205 nm.	94
Figura 3.50 Espectro medio de reflectancia SNV-D del suelo limoso con diésel a diferentes niveles de concentración entre la región 1087 a 1374 nm.	95
Figura 3.51 Segunda derivada de SNV-D del suelo limoso a diferentes niveles de concentración con diésel entre la región 1126 y 1264 nm.	96
Figura 3.52 Regresión lineal entre la concentraciones y la segunda derivada de la reflectancia SNV-D del suelo limoso con crudo de petróleo en la longitud de onda 1200 nm.	96
Figura 3.53 Espectro de reflectancia SNV-D del suelo arcilloso a diferentes niveles de concentración con crudo de petróleo entre la región 1087 a 1374 nm. . . .	97
Figura 3.54 Espectro de reflectancia SNV-D del suelo arcilloso a diferentes niveles de concentración con diésel entre la región 1087 a 1374 nm.	98
Figura 3.55 Segunda derivada de SNV-D del suelo arcilloso a diferentes niveles de concentración con diésel entre la región 1121 a 1259 nm.	98
Figura 3.56 Segunda derivada de SNV-D del suelo arcilloso a diferentes niveles de concentración con diésel entre la región 1126 a 1279 nm.	99
Figura 3.57 Regresión lineal entre la concentraciones y la segunda derivada de la reflectancia SNV-D del suelo arcilloso con crudo de petróleo en la longitud de onda 1224 nm.	100
Figura 3.58 Regresión lineal entre la concentraciones y la segunda derivada de la reflectancia SNV-D del suelo arcilloso con diésel en la longitud de onda 1200 nm.	100
Figura 3.59 Espectro de reflectancia SNV-D del suelo arenoso a diferentes niveles de concentración con crudo de petróleo entre la región 1087 a 1374 nm. . . .	101
Figura 3.60 Espectro de reflectancia SNV-D del suelo arenoso a diferentes niveles de concentración con diésel entre la región 1087 a 1374 nm.	102
Figura 3.61 Espectro de reflectancia de la segunda derivada SNV-D del suelo arenoso a diferentes niveles de concentración con crudo de petróleo entre la región 1131 y 1254 nm.	102
Figura 3.62 Espectro de reflectancia de la segunda derivada SNV-D del suelo arenoso a diferentes niveles de concentración con diésel entre la región 1126 y 1259 nm.	103

Figura 3.63	Regresión lineal entre las concentraciones y la segunda derivada de la reflectancia SNV-D del suelo arenoso con crudo de petróleo en la longitud de onda 1200 nm.	103
Figura 3.64	Regresión lineal entre las concentraciones y la segunda derivada de la reflectancia SNV-D del suelo arenoso con diésel en la longitud de onda 1200 nm.	104
Figura 3.65	Software Spectronon Pro.	112
Figura 3.66	Icono de Spectronon Pro.	113
Figura 3.67	Calibración del foco.	114
Figura 3.68	Calibración de la velocidad de la mesa.	114
Figura 3.69	Calibración de referencia oscura y blanca.	115
Figura 3.70	Hojas y flores.	115
Figura 3.71	Imagen tomada con la cámara Resono PIKA-NIR.	116
Figura 3.72	a) barra de herramienta, inicio de cámara b)utilidad del software para pasar de .BIL a .BSQ.	116

Índice de tablas

Tabla 1.1	Diferencia entre espectroscopia, análisis de imagen en visible e imágenes hiperespectrales (Martínez, 2015)	23
Tabla 1.2	Regiones del espectro electromagnético con sus respectivos intervalos de frecuencia, longitud de onda y energía (Hecht, 2000)	43
Tabla 3.1	Registros en el sensor (en mV) para I_0 e I_R	53
Tabla 3.2	Valores calculado de intensidad y error	54
Tabla 3.3	Características de la cámara Pika NIR-320	55
Tabla 3.4	Parametros del sistema hiperespectral	56
Tabla 3.5	Muestras de suelos y derivados de petróleo puros indicando algunas características	61
Tabla 3.6	Lista de mosaicos de los patrones de referencia	78
Tabla 3.7	Valores obtenidos de los tres tipos de suelos contaminados con crudo de petróleo.	86
Tabla 3.8	Valores obtenidos de los tres tipos de suelos contaminados con diésel.	87
Tabla 3.9	Espectros característicos de las muestras	92
Tabla 3.10	Grupos con sus respectivas características para el modelo lineal	104

Resumen

En esta investigación se estudiaron suelos contaminados con derivados de petróleo usando un sistema de imagen hiperespectral de laboratorio con el objetivo de analizar la respuesta espectral en la región de 900 a 1700 nm. Para ello se tienen como objetivos particulares preparar patrones de referencia de suelos derivados de petróleo, aplicar diferentes métodos de preprocesamientos y procesamientos para imágenes hiperespectrales y diferenciar el suelo contaminado del suelo no contaminado. El trabajo se divide en tres partes, (1) calibración del sistema de imágenes hiperespectrales de laboratorio, (2) preparación de las muestras patrones de referencia y adquisición de imágenes, (3) preprocesamiento y procesamiento de las imágenes hiperespectrales (HSI).

Para la calibración del sistema de imágenes hiperespectrales de laboratorio se debe realizar la calibración de la reflectancia de la cámara hiperespectral con un patrón de referencia (teflón) y la calibración del sistema HSI de laboratorio donde se controla la distancia foco-muestra, velocidad de la plataforma, fotogramas por segundo (fps) y resolución espacial.

Para realizar las muestras patrones se mezclan 70 % del suelo (limoso o arcilloso) con 30 % de crudo de petróleo, para luego reducir la concentración en un rango de 3 % hasta llegar a una concentración mínima de 3 % de crudo de petróleo, este procesamiento se repite para las muestras patrones con diésel. Para la muestra patrón de suelo arenoso la concentración máxima del contaminante es de 20 % (crudo o diésel) y se reduce en un rango de 2 % hasta llegar a una concentración mínima de 2 % (crudo o diésel). Se obtuvieron 60 patrones de referencia los cuales fueron comparadas los suelos sin contaminar (limoso, arcilloso y arenoso).

Para el preprocesamiento de las HSI se procedió con la normalización, se eliminó el ruido por medio de un suavizado, se corrigió el valor multiplicativo y la línea base de los valores obtenidos de las HSI, usando los métodos: normalización máxima, Savitzky-Golay, variable normal estándar (SNV) y de-trend.

Para el procesamiento de las HSI se utilizaron el análisis de componentes principales (PCA) y umbralización para poder eliminar los píxeles o clases (fondo, porta muestra, sombras o puntos de máxima reflectancia) no deseados de la muestra.

Después del procesamiento, se agruparon las imágenes por el tipo de contaminante para formar mosaicos y clasificarlo por PCA y analizar la dispersión de cada patrón de referencia frente a su valor medio (σ) y desviación estándar (s_n); logrando concluir que el valor medio de una muestra contaminada es mayor al valor medio de una muestra no contaminada, Observando que existe menor dispersión en las muestras contaminadas frente a la que no se contamina. Para poder relacionar la concentración con la segunda derivada de los espectros de reflectancia SNV-D, se realiza un modelo lineal entre las regiones de las longitudes de onda de 1087 a 1274 nm, obteniendo los mejores resultados

para el suelo limoso con crudo de petróleo. Con todos los resultados obtenidos en la investigación se concluyó que un sistema de imágenes hiperespectrales de laboratorio en el rango de 900 a 1700 nm se puede utilizar como un sistema rápido y no invasivo para clasificar entre suelo contaminado y no contaminado como también poder encontrar el grado de contaminación en los suelos contaminados por derivados de petróleo.

Abstract

In this investigation, soils contaminated with petroleum derivatives were studied using a laboratory hyperspectral imaging system with the objective of analyzing the spectral response in the region of 900 to 1700 nm. For this purpose, the main objectives are to prepare reference patterns of petroleum-derived soils, apply different methods of preprocessing and processing for hyperspectral images and differentiate contaminated soil from uncontaminated soil. The work is divided into three parts, (1) calibration of the laboratory hyperspectral imaging system, (2) preparation of reference standards samples and image acquisition, (3) preprocessing and processing of hyperspectral images (HSI).

For the calibration of the laboratory hyperspectral imaging system the calibration of the reflectance of the hyperspectral chamber with a reference standard (Teflon) and the calibration of the laboratory HSI system where the focus-sample distance, speed of the sample should be controlled platform, frames per second (fps) and spatial resolution.

To make the standard samples, 70 % of the soil (silty or clayey) is mixed with 30 % of crude oil, then reduce the concentration in a range of 3 % until reaching a minimum concentration of 3 % of crude of oil, this processing is repeated for standard samples with diesel. For the standard sample of sandy soil, the maximum concentration of the contaminant is 20 % (crude or diesel) and is reduced in a range of 2 % until reaching a minimum concentration of 2 % (crude or diesel). 60 reference standards were obtained which were compared to uncontaminated soils (silty, clayey and sandy).

For the preprocessing of the HSI, normalization was carried out, the noise was eliminated by means of smoothing, the multiplicative value and the baseline of the values obtained from the HSI were corrected, using the methods: maximum normalization, Savitzky-Golay, standard normal variable (SNV) and de-trend.

For the HSI processing, principal component analysis (PCA) and thresholding were used to eliminate unwanted pixels or classes (background, sample holder, shadows or points of maximum reflectance) from the sample.

After processing, the images were grouped by the type of contaminant to form mosaics and classify it by PCA and analyze the dispersion of each reference standard against its average value (*sigma*) and standard deviation (s_n); managing to conclude that the average value of a contaminated sample is greater than the average value of an uncontaminated sample, noting that there is less dispersion in the contaminated samples compared to the one that is not contaminated. In order to be able to relate the concentration to the second derivative of the SNV-D reflectance spectra, a linear model is made between the wavelength regions of 1087 to 1274 nm, obtaining the best results for silty soil with petroleum oil. With all the results obtained in the investigation,

it was concluded that a system of laboratory hyperspectral images in the range of 900 to 1700 nm can be used as a rapid and non-invasive system to classify between contaminated and uncontaminated soil as well as to find the degree of contamination in soils contaminated by petroleum derivatives.

INTRODUCCIÓN

Las imágenes hiperespectrales han sido aplicadas desde un inicio en la teledetección para luego expandirse en otras áreas como: investigación farmacéutica (Martínez, 2015), control de calidad de los alimentos (Park and Lu, 2015), ciencias forenses, bioquímica, biomedicina (Lu and Fei, 2014), medio ambiente (Amigo et al., 2015), prospección minera, entre otros (Reséndez-Hernández et al., 2018). Las aplicaciones de imágenes hiperespectrales se realiza en los sistemas de laboratorio, aire libre y aerotransportado, además se puede emplear en las regiones espectrales del infrarrojo cercano, infrarrojo medio, ultravioleta y visible. Las imágenes hiperespectrales que se adquieren tienen la ventaja de combinar imágenes tradicionales y espectroscopia haciendo que por cada píxel de la imagen se obtenga un espectro de cientos de bandas o longitud de onda. La información que se obtiene de una imagen hiperespectral se distribuye en un hipercubo en el cual los ejes x , y es la posición del píxel y el eje z es la banda (Dai et al., 2014; Amigo et al., 2015). El análisis hiperespectral está inherentemente vinculado al análisis de datos multivariado y a la implementación de diferentes algoritmos para el manejo de todos los datos generados por una imagen hiperespectral.

Debido a su gran potencial las imágenes hiperespectrales se vienen aplicando recientemente en investigaciones sobre contaminación de suelos con derivados de petróleo. Los suelos contaminados con petróleo usualmente han sido analizados por medio de *espectroradiómetros* en el visible e infrarrojo cercano (Chakraborty et al., 2010; Schwartz et al., 2013; Reséndez-Hernández et al., 2018; García et al., 2019). Otras investigaciones utilizan sistemas aerotransportados con cámaras multiespectrales para el análisis de suelos contaminados (Kühn et al., 2004), mientras que otras combinan el uso de los espectroradiómetros con imágenes satelitales para detección de suelo contaminado con petróleo (Andreoli et al., 2007). Recientes investigaciones utilizan imágenes hiperespectrales y análisis multivariable para la clasificación y cuantificación de contaminantes en suelos (El-Rahman and Zolait, 2018), (Pelta and Ben-Dor, 2019), (Fan et al., 2019). Así mismo para estudiar las propiedades del petróleo usando análisis de componente principales (Long et al., 2019), así como para el análisis de contaminantes de forma indirecta en vegetación (Lassalle et al., 2018). Todos los trabajos analizan la absorbancia, la reflectancia difusa, primera derivada o segunda derivada. Se llega a la conclusión que el uso de las imágenes hiperespectrales en el infrarrojo o visible pueden detectar los suelos contaminados con dichos derivados, pudiendo cuantificar la concentración y clasificar los contaminantes en los suelos.

Teniendo en cuenta los antecedentes mencionados, la presente investigación está dirigida a realizar mediciones con un sistema de imagen hiperespectral de laboratorio para tres diferentes suelos contaminados con derivados de petróleo con diferentes

concentraciones con el objetivo de cuantificar y clasificar los tipos de contaminación, utilizando la respuesta de reflectancia, la primera y segunda derivada en la región de 900 a 1700 nm, la cual no se ha desarrollado antes de este trabajo.

Planteamiento del problema

El suelo es un medio en el que ocurren una gran cantidad de procesos naturales y su calidad está determinada por su capacidad para sustentar la vida vegetal y animal (Guide, 1999); sin embargo está expuesto a diferentes fuentes de contaminación como los derrames de derivados de petróleo, el cual tiene efectos adversos para la naturaleza, modificando la estructura y el funcionamiento de los suelos.

La gran tasa de producción de 100.8 millones de barriles de petróleo crudo por día (AIE, 2019) y el uso generalizado de derivados de petróleo para el transporte y la industria ha llevado a derrames accidentales, fugas a largo plazo o fallas operacionales. En consecuencia, la contaminación por derivados de petróleo y sus efectos pueden encontrarse en todo el mundo. La manera de identificar la concentración y clasificar estos derrames en suelo se realiza por medio de diferentes técnicas como: espectrometría, imágenes satelitales y recientemente sistemas hiperspectrales (Reséndez-Hernández et al., 2018). Esta última técnica combina el análisis de imagen y la espectrometría para identificar y cuantificar la muestra de manera rápida sin manipularla y destruirla. Sus aplicaciones van desde la teledetección, la medicina, la agricultura y recientemente en contaminantes.

El Perú no está excepto de los casos de derrame de petróleo debido que se producen alrededor de 36 millones de barriles anuales de petróleo crudo. La selva es la región más productiva, con un promedio de 24.8 millones de barriles al año (equivalente a 68000 barriles por día), la costa le sigue con 6.8 millones de barriles al año (18500 barriles diarios) y el resto es extraído del zócalo continental por medio de plataformas en el mar (Minagri, 2019). En Perú, el estándar de calidad ambiental (ECA) para suelo, publicado en el Decreto supremo N^o 011-2017-MINAM indica que, para el análisis de muestras de suelo con hidrocarburos de petróleo se aplica el EPA 8015, la cual indica el uso de cromatografía de gases, siendo aplicable para un número reducido de muestras en forma puntual.

Una alternativa propuesta para medir estos contaminantes en el suelo es el uso de la técnica de imágenes hiperspectrales, debido a que analiza grandes cantidades de muestras en forma continua y la posibilidad de utilizar en el mismo lugar del derrame. En el laboratorio de investigación de la Facultad de Ciencias en la Universidad de Ingeniería se cuenta con el único sistema de imágenes hiperspectrales de 900 a 1700 nm en el país. Dentro de esta región existen bandas características que relacionan los derivados de petróleo en el suelo; pero no se han analizado con mayor detalle.

La pregunta de investigación del presente trabajo es: ¿Usando HSI en el rango de 900 a 1700 nm existirá alguna respuesta espectral y espacial en los suelos contaminados con derivados del petróleo?

Hipótesis

Utilizando HSI en el rango NIR de 900 a 1700 nm junto con métodos multivariados se logra clasificar los suelos contaminados con derivado de petróleo y conocer su concentración de contaminación

Justificación

La contaminación del suelo es una amenaza latente producida por el desarrollo económico y social de las comunidades. La actividad petrolera es una de las industrias que genera mayor impacto ambiental. Los derrames de hidrocarburos de petróleo son comunes durante los procesos de producción, comercialización, transporte y almacenamiento.

Mantener un suelo sano garantiza la continuidad de los procesos naturales. Los estudios basados en sistemas hiperespectrales se han llevado a cabo principalmente en suelos afectados por derrames de hidrocarburos, donde se analizaron varios casos de filtración de hidrocarburos y se llegó a la conclusión que la tecnología hiperespectral puede detectar el petróleo presente y tiene una gran capacidad de cuantificar y caracterizar el contenido de hidrocarburos (Asadzadeh and de Souza Filho, 2016).

Para interpretar correctamente los datos hiperespectrales, los espectros de reflectancia recuperados deben estar correlacionadas con materiales específicos; Por lo tanto, se deben construir bibliotecas espectrales que contengan los espectros de reflectancia de los materiales a detectar. Esto requiere mediciones de luz reflejada de las muestras del material investigado que se realizan en el laboratorio o en el campo.

En el Perú no existe información estadística oficial sobre la ubicación y extensión de los suelos contaminados con hidrocarburos. Esto representa un problema que se debe afrontar para poder evitar la contaminación de las aguas subterráneas.

Las imágenes hiperespectrales brindan amplia información para identificar y distinguir entre materiales espectralmente similares (pero únicos); lo que brinda la capacidad de hacer distinciones adecuadas entre materiales con espectros de reflectancia ligeramente diferentes. Por lo expuesto se justifica una investigación para evaluar la metodología para identificar los suelos contaminados con petróleo.

Objetivos

Objetivos generales

- Analizar la respuesta espectral en la región de 900 a 1700 nm de diferentes suelos contaminado con derivados de petróleo.

Objetivos particulares

- Preparar las muestras patrones de diferentes suelos con cada uno de los derivados de petróleo a estudiar.
- Aplicar los diferentes métodos de preprocesamiento y procesamiento en las imágenes hiperespectrales obtenidas.
- Clasificar y cuantificar la presencia de derivados de petróleo en los suelos.

Metodología

Esta investigación es cuantitativa, ya que sigue un orden durante el proceso de investigación y utiliza métodos matemáticos y estadísticos para extraer conclusiones.

El alcance de este estudio es descriptivo y correlacional, ya que consiste en describir el nivel de contaminación de las muestras a partir de las imágenes hiperespectrales y es correlacional debido que relaciona la reflectancia de algunas bandas entre 900 a 1700 nm y la concentración del suelo contaminado.

El diseño de la investigación es experimental, ya que se manipula el nivel de concentración de los derivados de petróleo (crudo, diésel y gasolina) y el tipo de suelo (limoso, arcilloso y arenoso).

La metodología de la presente investigación se menciona a continuación. Para preparar los patrones de referencia de suelos contaminados, se secan los suelos (limoso, arcilloso y arenoso) hasta eliminar la máxima cantidad de humedad y luego se tamiza para obtener los suelos de forma homogénea, esta metodología se menciona en los trabajos de Martínez et al. (2001); Andreoli et al. (2007). Para realizar las mezclas entre los suelos y derivados de petróleo como lo sugiere Martínez et al. (2001); Andreoli et al. (2007); Pelta and Ben-Dor (2019) se realiza una combinación del 70 % de suelo arcilloso o suelo limoso con 30 % de crudo de petróleo ligero o diésel de manera respectiva para obtener cuatro patrones de referencia de máxima concentración. En el caso del suelo arenoso se realiza la combinación de 80 % de suelo arenoso con 20 % de crudo de petróleo ligero o diésel respectivamente. Estos seis patrones de referencia con máxima concentración se reducen para obtener 10 niveles de concentración para cada uno de los seis patrones de referencia inicial (Pelta and Ben-Dor, 2019).

La metodología que se utiliza en la calibración, se encuentra en Resonon (2017), en ella se indica los parámetros que se tiene que tomar en la calibración. Para el sistema de cámara hiperespectral son: distancia de la muestra a la cámara y tiempo de integración. Para la calibración de la cámara hiperespectral son: resolución espacial y resolución espectral (en la calibración de la reflectancia de la cámara hiperespectral se utiliza el hipercubo de referencia del teflón y un hipercubo de la corriente oscura, los cuales se adquieren con la cámara hiperespectral).

En la adquisición de las imágenes hiperespectrales se utiliza el software Spectronon-Pro (Resonon, 2017), el cual adquiere las imágenes en el modo de detección de reflectancia y modo de escaneo de banda intercalada por línea.

En el preprocesamiento se utilizan los siguientes métodos junto con las librerías del lenguaje de programación R para los valores obtenidos de las imágenes hiperespectrales: reducción de dimensiones espectrales (Park and Lu, 2015), normalización máxima (Cao et al., 2017), suavizado Savitzky Golay (Savitzky and Golay, 1964), corrección de escala multiplicativa SNV y corrección de línea base de-trend Park and Lu (2015), segunda derivada espectral (Manolakis et al., 2003) y análisis de componente principales PCA Manolakis et al. (2003); Park and Lu (2015); Bajorski (2011).

Se realizan diferentes procesamientos en los valores de las imágenes hiperepectrales para: obtener la librería de los espectros de reflectancia de los suelos contaminados con derivados de petróleo, clasificar un suelo contaminado y no contaminado y para relacionar los diferentes niveles de concentraciones con la reflectancia de una longitud de onda. El esquema de la metodología realizada en el presente trabajo se puede observar en la figura 1.12.

CAPÍTULO 1

1. Imágenes hiperespectrales, adquisición y procesamiento

1.1. Imágenes hiperespectrales (HSI)

Una imagen digital es una representación numérica en 2 dimensiones de un objeto. Las imágenes digitales generalmente se refieren a imágenes **raster** (archivos que contienen elementos discretos llamados píxeles, siendo el píxel la menor unidad que forma parte de una imagen digital). Una imagen hiperespectral es el conjunto secuencial de cientos de imágenes digitales a cada imagen digital le corresponde una banda la cual está asociada a una longitud de onda continua en una región espectral (ultravioleta, visible o infrarrojo) donde a cada banda se le puede asociar una medición física, química o biológica (Park and Lu, 2015; Amigo et al., 2015). En la figura 1.1 se muestra una imagen hiperespectral de píxeles de x columnas e y filas también llamados muestras, líneas y λ longitudes de ondas. Se observa que un píxel de la imagen hiperespectral contiene un espectro continuo de longitud de onda con su respectiva intensidad.

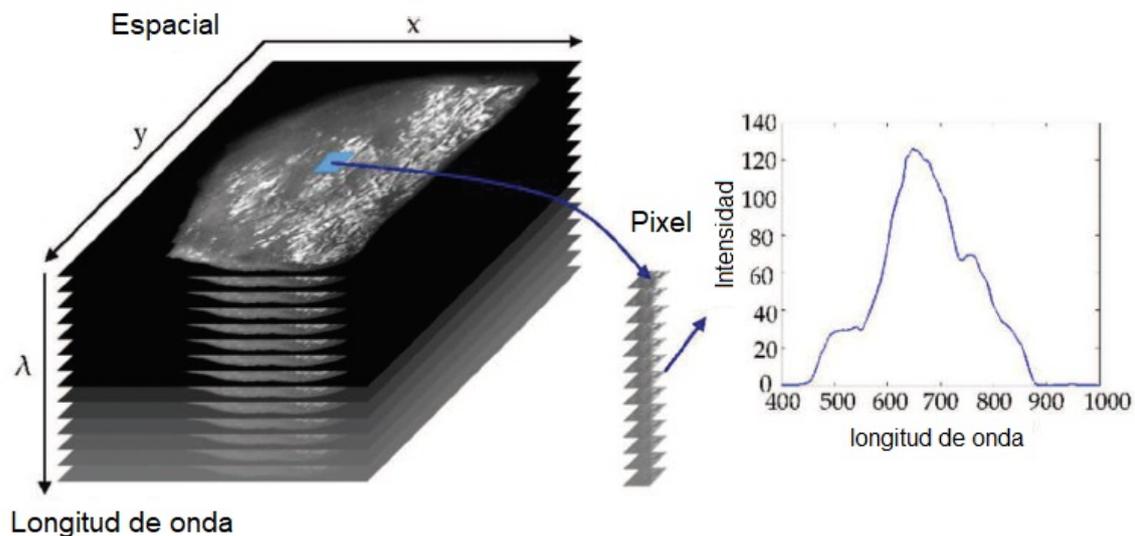


Figura 1.1: Imagen hiperespectral (Wang and Peng, 2018).

El uso de imágenes hiperespectrales presenta diferentes ventajas frente a otras técnicas convencionales. En la tabla 1.1 se comparan la técnica de imágenes tradicionales (RGB), la espectroscopía y las imágenes hiperespectrales. La espectroscopía no presenta información espacial. Las imágenes tradicionales (RGB) no presentan información espectral como tampoco pueden medir características físicas, químicas o biológicas de

una muestra. Las imágenes hiperespectrales tienen información espacial, espectral e información física, química y biológica.

Tabla 1.1: Diferencia entre espectroscopia, análisis de imagen en visible e imágenes hiperespectrales (Martínez, 2015)

Características	Imágenes tradicionales	Espectroscopia	Hiperespectral
Información espacial	Sí	No	Sí
Información espectral	No	Sí	Sí
Física, química y biológica	No	Sí	Sí

1.2. Hipercubo

El hipercubo es un arreglo tridimensional de valores de reflectancia, absorbancia o transmitancia según sea el modo de detección utilizado. Estos valores están asociados a cada píxel y banda de la imagen hiperespectral. En la figura 1.2 se observan los hipercubos de una imagen multispectral y de una imagen hiperespectral, en sus coordenadas x e y se ubican los píxeles de una longitud de onda λ correspondiente a una determinada banda. Las bandas se denominan 1, 2 y así sucesivamente hasta completar todas las bandas que pudiera tener el hipercubo. La diferencia entre el hipercubo de una imagen multispectral y el de una imagen hiperespectral es la cantidad de bandas. Un hipercubo se redimensiona en una matriz la cual llamamos *matriz aleatoria* de n filas y p columnas, en la matriz aleatoria que se observa en la figura 1.2, el valor X_{11} es el primer píxel de la banda 1 y el valor X_{n1} es el último píxel de la banda 1, mientras que el valor X_{1p} es el primer píxel de la banda p y el valor X_{np} es el último píxel de la banda p . Para el presente trabajo se tienen 168 bandas.

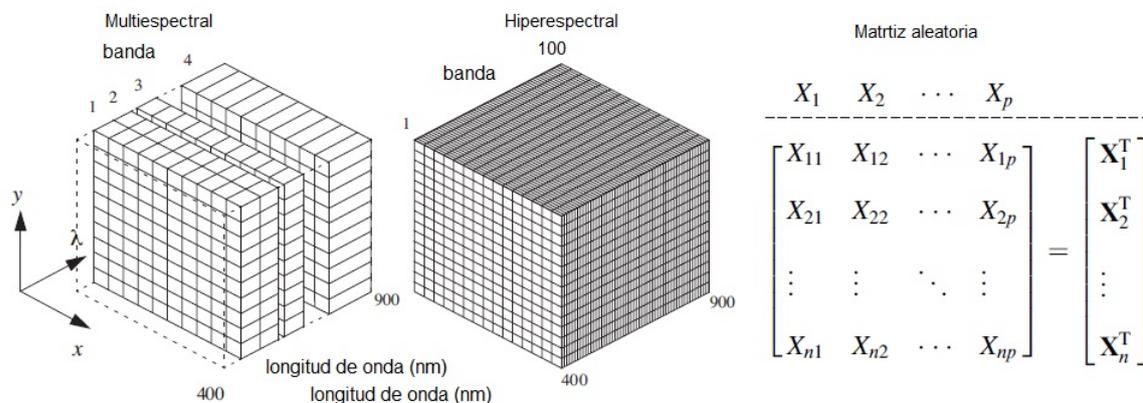


Figura 1.2: Hipercubo de una imagen multispectral, imagen hiperespectral y la matriz aleatoria. Adaptado de Falco (2015); Bajorski (2011).

La matriz aleatoria se puede representar en una forma vectorial $\mathbf{X} = [\vec{X}_1, \vec{X}_2, \dots, \vec{X}_p]$,

donde los valores $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_p$ son los vectores para cada banda que contiene todos los píxeles de las respectivas bandas (Falco, 2015). También la matriz aleatoria \mathbf{X} se puede describir como el vector de n píxeles $\vec{X}_1^T; \vec{X}_2^T; \vec{X}_3^T; \dots; \vec{X}_n^T$ que contienen las p bandas (Bajorski, 2011).

1.3. Tipos de sistemas de adquisición

Los sistemas de adquisición de imágenes hiperespectrales son variados y dependen de la aplicación o investigación que se realice. Se clasifican en sistemas de laboratorios, al aire libre y aerotransportado, cada una presenta características y ventajas frente a otras.

1.3.1. Sistema de imagen hiperespectral de laboratorio

El sistema hiperespectral de laboratorio está conformado por diferentes partes, las cuales son: cámara hiperespectral, lente, torre, fuentes de luces (lámparas de tungsteno), plataforma, motor y una computadora (ver figura 1.3). La característica más relevante de un sistema de adquisición de laboratorio es poder controlar los parámetros de intensidad de luz, distancia entre la lente y cámara, los fotogramas por segundos y el movimiento de la plataforma (Yin et al., 2017; Zhu et al., 2017; Feng et al., 2018).

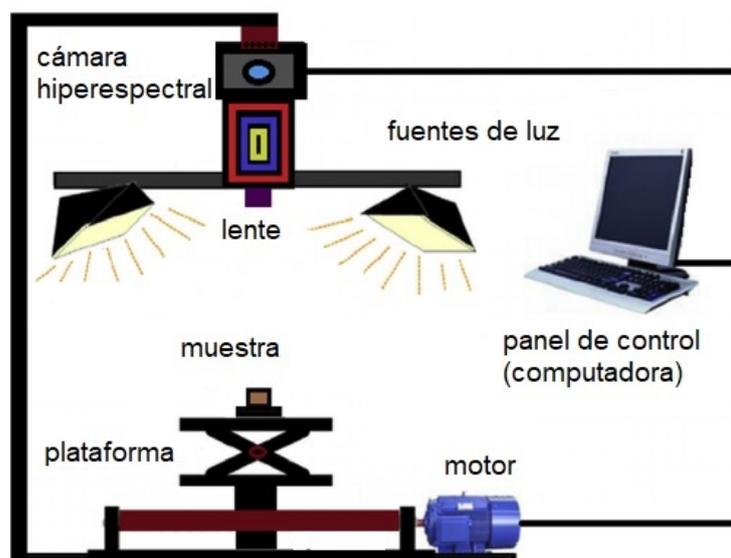


Figura 1.3: Sistema de imagen hiperespectral de laboratorio donde se nombra las partes que lo caracterizan (cámara hiperespectral, lente, plataforma, muestra, motor, luces y computadora). Adaptado de Mishra et al. (2017).

1.3.2. Sistema al aire libre

En un sistema hiperespectral al aire libre se trabaja con fuente de luz natural o artificial. En cualquier caso se necesita un objetivo de referencia (teflón o spectralon) para la calibración, la cual se compara con el objetivo de reflectancia tomado en el laboratorio a condiciones controladas (fuente de luz). En este sistema no es necesario una plataforma móvil debido que el sistema toma la imagen hiperespectral en modo de escaneo de banda secuencial (BSQ), (Catelli et al., 2018; Pelta and Ben-Dor, 2019; Park and Lu, 2015). En la figura 1.4 se tiene un ejemplo del sistema hiperespectral al aire, el cual capta la reflectancia de la planta al incidir la luz natural sobre ella.

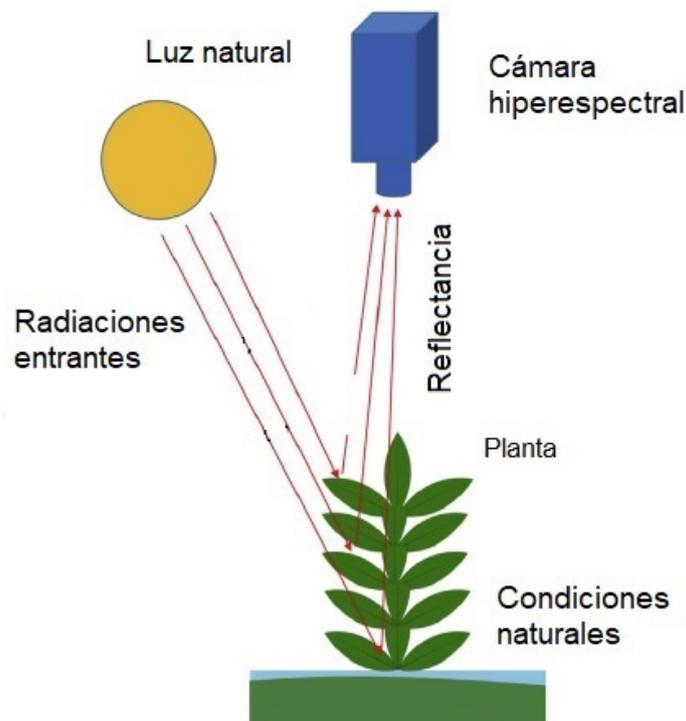


Figura 1.4: Sistema al aire libre con fuente de luz natural. Adaptado de Mishra et al. (2017).

1.3.3. Sistema aerotransportado

Un sistema aerotransportado nos brinda información de un área u objeto sobre el cual sobrevuela el sistema por medio de un Vehículo aéreo no tripulado (UAV por las siglas en inglés *Unmanned Aerial Vehicle*). Se divide en dos categorías: activos y pasivos (figura 1.5). Ambos miden la energía reflejada de un área. Los sensores activos producen su propia luz (luz solar) la cual se proyecta, refleja y detecta. La medición pasiva utiliza una fuente de luz natural. La corrección de los datos hiperespectrales en el aire es importante para eliminar los efectos de las condiciones atmosféricas, de lo

contrario puede llevar a una interpretación de datos incorrectos (Mishra et al., 2017; Correal and García, 2012).

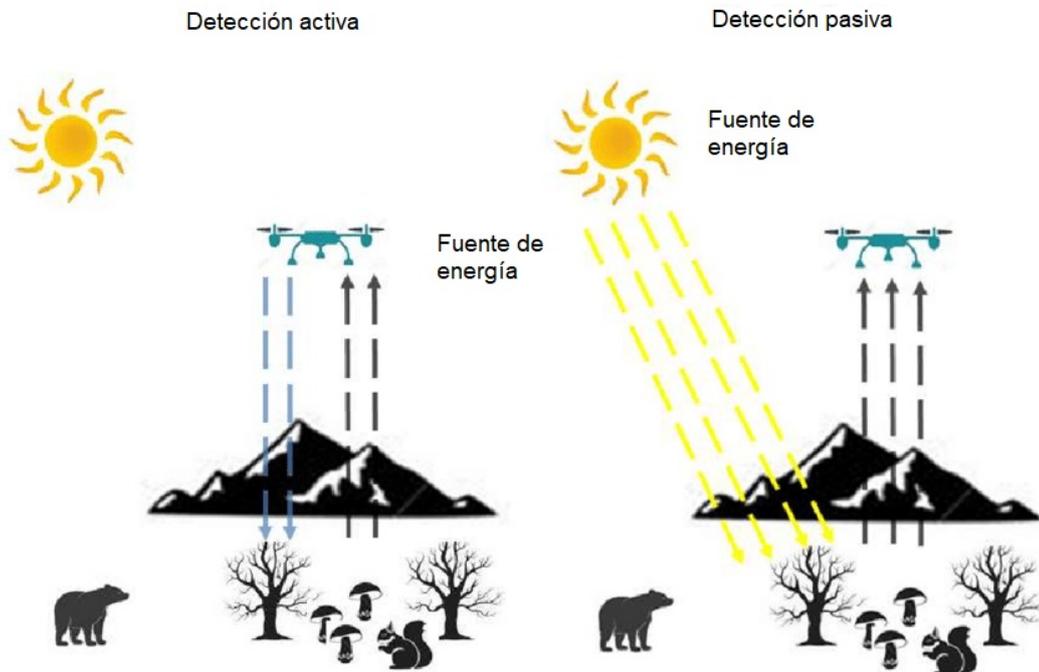


Figura 1.5: Sistema aerotransportado activo y pasivo (Analytik, 2019).

1.4. Modo de escaneo y de detección

Para la adquisición de una imagen hiperespectral existen tres modos de escaneo y tres modos de detección.

Los modos de escaneo son: banda intercalada por línea (BIL), banda secuencial (BSQ) y banda intercalada por píxel (BIP). Los modos de detección son: reflectancia, transmitancia e interactancia.

1.4.1. Modo de escaneo de una cámara hiperespectral

Los modos de escaneo se refieren al modo que se obtienen los valores de píxeles de cada banda para formar el hipercubo. Estos modos son: La banda intercalada por línea (BIL) y consiste en obtener los valores de píxeles de una línea y todas las bandas (ver A en la figura 1.6), la banda secuencial (BSQ) y consiste en todos los píxeles de una sola banda (ver B en la figura 1.6); la banda intercalada por píxel (BIP) y consiste en un solo píxel y todas las bandas (ver C en la figura 1.6). Los modos BIL, BIP y BSQ son archivos binarios y debe tener un archivo encabezado para que puedan ser interpretados correctamente los valores que forman el hipercubo obtenidos por alguno de estos modos. El archivo de encabezado contiene datos secundarios sobre el número

de píxeles, bandas y otros que permitan el ordenamiento del hipercubo y sus valores pertenecientes a los datos primarios.

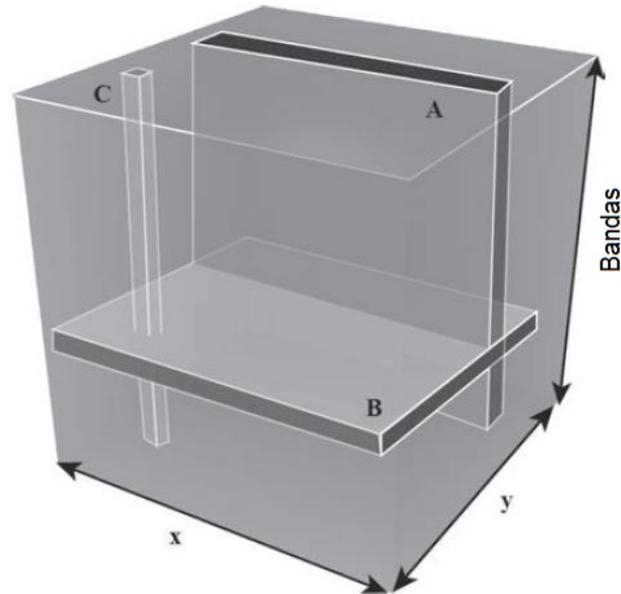


Figura 1.6: Cubo de datos que muestra la dimensión espacial (x,y) y la dimensión espectral (λ). Las regiones, etiquetadas A, B y C, ilustran la porción del cubo de datos que se adquiere durante un ciclo de adquisición (Li et al., 2018a).

El archivo BIL es el más común para un sistema de imagen hiperespectral NIR. La figura 1.7 muestra el modo de escaneo por barrido y como se forma el archivo BIL. En el modo de escaneo por barrido la mesa tiene una dirección de movimiento. Los elementos de una superficie que se adquieren son: píxeles, franja de la muestra y proyección de formación del detector. El campo de visión (FOV por las siglas en inglés *Field of View*) y campo de visión instantánea (IFOV por las siglas en inglés *Instantaneous Field of View*) es dado por el tipo de lente de la cámara.

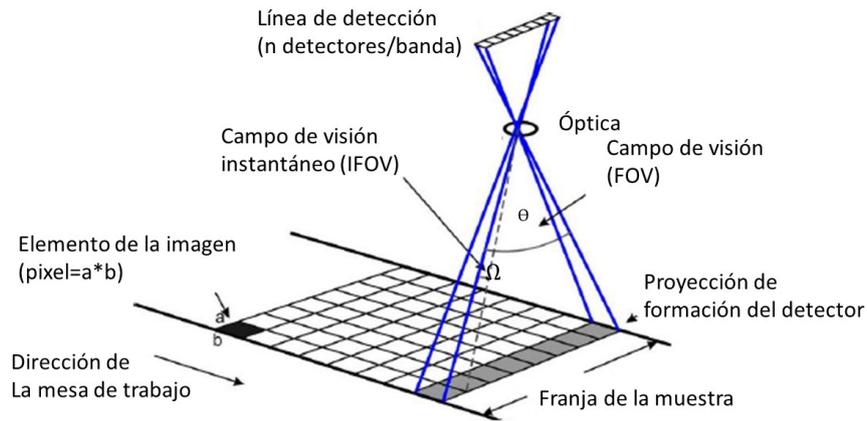


Figura 1.7: Escáner de línea de barrido. Adaptado de Manolakis et al. (2016).

El tamaño del píxel ($a*b$) dependerá del ángulo solido asociado (Ω) al IFOV y del ángulo (θ) asociado al FOV. La ecuación 1 es utilizada para calcular el tamaño del píxel, en ella se observa la altura de la lente óptica a la franja de la muestra (h), el ángulo asociado al FOV y los píxeles contenidos en la franja (N).

$$b = a = \frac{2 * h * \tan(\theta/2)}{N} \quad (1)$$

1.4.2. Modo de detección de una cámara hiperespectral

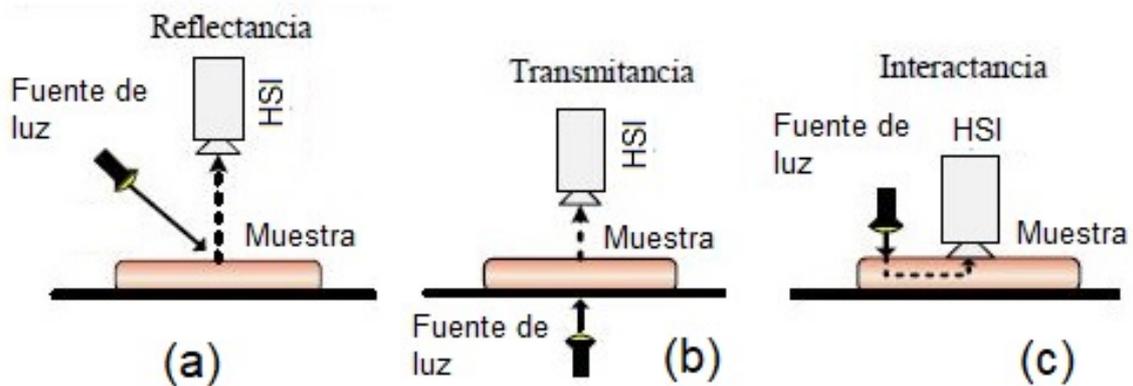


Figura 1.8: Modos de detección de una cámara hiperespectral. Adaptado de Castro (2015).

Los modos de detección de un sistema de imágenes hiperespectrales son: reflectancia, transmitancia e interactancia. El sensor en el modo de reflectancia captura la luz reflejada de forma que la cámara hiperespectral esté frente a la muestra. Este modo es

utilizado para la obtención de características de calidad externas como la forma, tamaño, textura de la superficie, entre otras (figura 1.8 a). El modo de transmitancia es cuando el sensor captura la luz transmitida a través de la muestra y lleva información interna, este modo es utilizado generalmente para determinar la concentración de los componentes y defectos internos de los objetos transparentes (figura 1.8 b). El modo de interactancia es cuando la fuente de luz y el sensor están situados en el mismo lado de la muestra de forma paralela (figura 1.8 c) (Gowen et al., 2007; Manolakis et al., 2016; Castro, 2015). En el presente trabajo se utiliza el modo de reflectancia.

1.4.3. Componentes de una cámara hiperespectral

En la figura 1.9 observamos los componentes que conforman un típico sistema de imagen hiperespectral y cómo el hipercubo de datos se arma uniendo cada línea escaneada de una muestra, mientras la muestra se mueve horizontalmente a una velocidad adecuada. Los componentes que conforman una cámara hiperespectral son: lente de enfoque, abertura, lente colimador, elemento dispersivo y lente de la cámara. El área de la muestra que se escanea recorre los componentes mencionados para poder llegar al dispositivo de carga acoplada (CCD). El componente clave de la cámara hiperespectral es el dispositivo de dispersión de longitud de onda, los cuales pueden ser: filtro, rejilla o prisma. Estos dispositivos ópticos se utilizan para dispersar la luz de banda ancha en diferentes longitudes de onda y proyectar la luz dispersada en el detector de área. Los filtros siempre se usan en el sistema de imágenes multispectrales, mientras que el prisma y la rejilla se usan ampliamente en el sistema de imágenes hiperespectrales (Zhang et al., 2014).

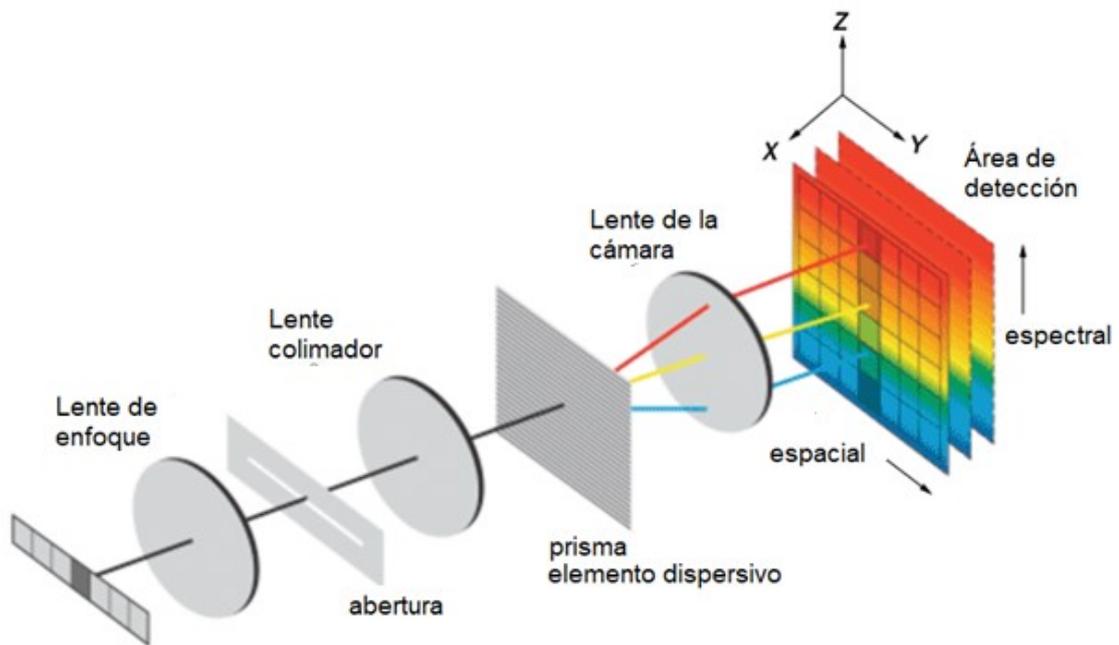


Figura 1.9: Componentes de una cámara hiperespectral para conformar un hipercono. Adaptado de Manolakis et al. (2016).

La cámara es una componente central del sistema de imágenes hiperespectrales. El dispositivo de carga acoplada (CCD) y los sensores de imagen de semiconductores de óxido metálico complementario (CMOS) son dos medios diferentes para generar la imagen digitalmente (Zhang et al., 2014), ambos sensores están formados por semiconductores de óxido metálico (MOS) distribuidos en forma de celdas en una matriz que convierten las cargas de las celdas en voltajes para entregar una señal. La diferencia principal entre el CCD y CMOS es: en el CCD la señal analógica se entrega en la salida y es digitalizada por la cámara, mientras que en el CMOS se realiza la digitalización de los píxeles internamente en unos transistores que lleva cada celda. En general, el sensor de imagen CMOS se usa en aplicaciones de calidad menos exigentes y el sensor de imagen CCD se usa en aplicaciones médicas, científicas y profesionales donde se requieren datos de imagen de alta calidad.

1.5. Calibración

La calibración de los datos de una imagen hiperespectral es un punto importante para garantizar la precisión y reproducibilidad del experimento e incluye: calibración espectral, espacial y radiométrica.

La calibración espectral vincula los números de banda con las longitudes de onda. La calibración espacial (o geométrica) es un proceso que correlaciona cada píxel de imagen con unidades como patrones de cuadrícula y proporciona información sobre

las dimensiones espaciales de cada píxel del sensor en la superficie del material, la ubicación y el tamaño de la muestra. La calibración radiométrica fuera de laboratorio se refiere a la conversión de los números digitales (datos sin procesar) a unidades físicas de luminosidad para luego pasarlo a reflectancia, en aplicaciones de laboratorio la calibración radiométrica consiste en calibrar la reflectancia directamente de un material estándar que son espectralmente planos y espacialmente homogéneos. Debido a que se utiliza un material de superficie plana, también se utiliza el término de calibración (o corrección) de campo plano.

1.6. Preprocesamiento espacial

Es un proceso para mejorar o manipular la información espacial de la imagen. Se pueden aplicar las técnicas de procesamiento de imágenes convencionales, excepto el preprocesamiento espacial para eliminar ruido; ya que dicho preprocesamiento espacial puede afectar a los espectros de reflectancia. Algunos métodos importantes de preprocesamiento espacial son: segmentación de imagen, enmascaramiento e imagen mosaico.

1.6.1. Segmentación de imagen

El preprocesamiento espacial en imágenes hiperespectrales se inicia determinando la ubicación espacial de la muestra que se analiza (conocer la posición de los píxeles contenidos en la muestra), este proceso generalmente comienza segmentando (separando) la imagen por medio de un umbral (intensidades en la que están los píxeles de interés) aplicado al histograma de las intensidades de los píxeles de la imagen en una banda seleccionada. Un método de segmentación es la binarización de la imagen, esto se puede ver en la figura 1.10, la cual produce una imagen binaria del fondo (color blanco) y de la zona que se analiza (color negro). Esta imagen binaria está asociada a un histograma al cual se puede realizar una umbralización sólo para la máscara binaria que se encuentra entre los valores de -3 y -1.5.

A partir de una imagen segmentada se puede realizar una máscara, la cual consiste en ubicar la posición de los píxeles de interés de la imagen segmentada para luego ubicar las posiciones en una imagen hiperespectral (figura 1.10).

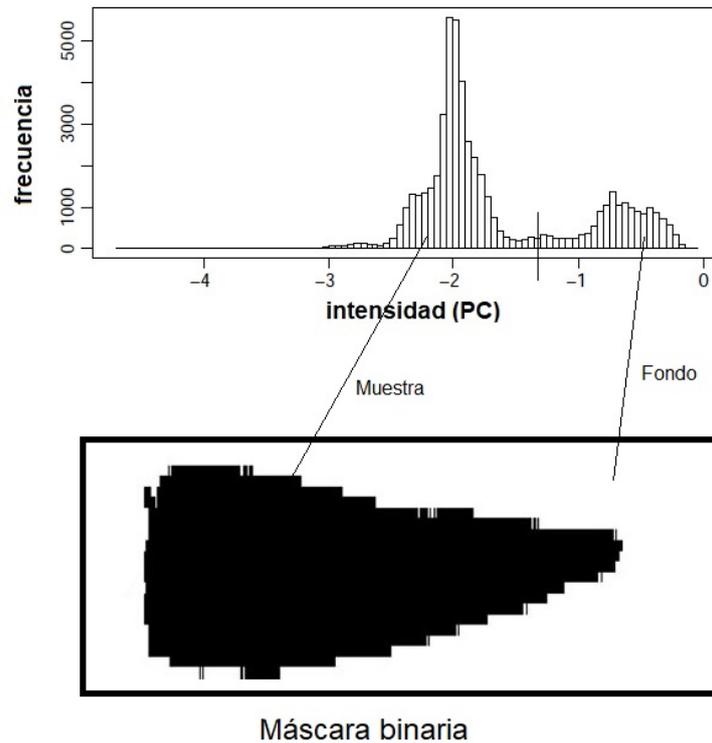


Figura 1.10: Umbralización por medio de un histograma y enmascaramiento binario (Bajorski, 2011).

1.6.2. Región de interés (ROIs)

Las regiones de interés (ROIs) se refiere a la selección de un subconjunto de muestras para extrapolar la población de las muestras seleccionadas (Yoon et al., 2013). Los ROIs representan la población de cada material, si hay varios materiales para examinar, los ROIs de todos los materiales se preparan de tal manera que los píxeles que muestran espectros puros conformen cada ROI. Los destellos y las sombras a menudo se excluyen en el ROI. Los píxeles a lo largo de los límites de cada material tampoco se incluyen en el ROI. Los ROIs se utilizan para construir bibliotecas espectrales, diseñar modelos de clasificación y evaluar el rendimiento de los modelos de clasificación. (Park and Lu, 2015; Bajorski, 2011).

1.6.3. Mosaicos

La representación de imágenes basada en mosaicos se puede utilizar para facilitar el análisis de datos y el desarrollo del algoritmo de clasificación, donde un mosaico de múltiples imágenes hiperespectrales se fusiona en una sola imagen hiperespectral (Yoon et al., 2013). El mosaico de imágenes puede acelerar el análisis de la imagen, el desarrollo de la clasificación y la evaluación utilizando un solo cubo de datos hiperespectrales (Park and Lu, 2015).

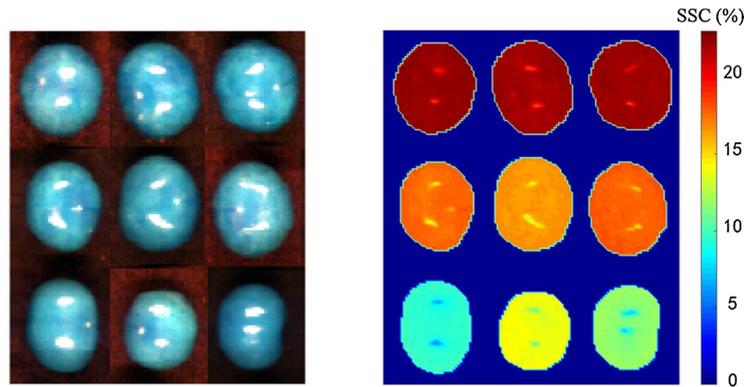


Figura 1.11: Mosaico de una imagen hiperespectral de cerezas (Li et al., 2018b).

En la figura 1.11 se tiene un mosaico conformado por diferentes imágenes hiperespectrales de cerezas con un mismo fondo. Al procesar la imagen hiperespectral del mosaico y seleccionar una banda, podemos encontrar las diferencias de las nueve cerezas en una escala de colores de 0 a 20% de SSC.

1.7. Preprocesamiento espectral

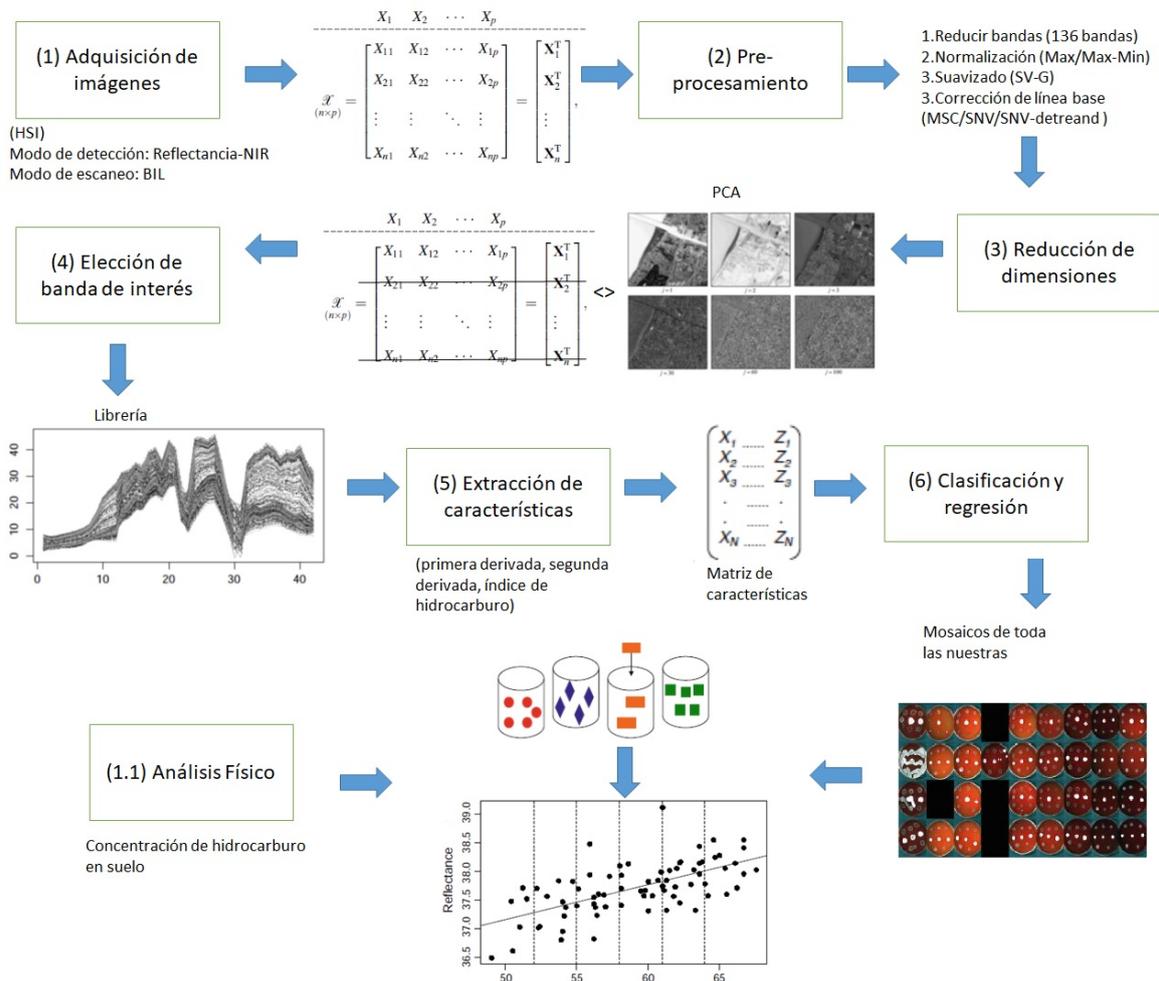


Figura 1.12: Metodología en el análisis de imágenes hiperespectrales (HSI) para suelos contaminados con derivado de petróleo (DP).

El preprocesamiento espectral en las imágenes hiperespectrales se agrupan en dos categorías: la primera es la extracción de miembros finales, donde un miembro final se refiere a un espectro de reflectancia. Los miembros finales se utilizan para construir una biblioteca espectral de espectros puros (espectros de muestras sin alterar, ejemplo: el agua, el suelo, la vegetación, etc). Estos miembros finales se extraen de una imagen dada, generalmente se obtienen mediante métodos de preprocesamiento espectral como el índice de pureza de píxeles (PPI). La segunda categoría es el uso de métodos estadísticos (ejemplo: análisis multivariado) y matemáticos en los espectros de absorción, reflectancia, etc. Los valores espectrales obtenidos mediante el preprocesamiento espectral se utilizan normalmente en el análisis multivariado, como PCA y mínimos cuadrados parciales (PLS) (Ozaki et al., 2006; Williams et al., 1987). Por lo tanto, la selección de métodos de preprocesamiento depende de los datos obtenidos y de la tarea

que se busca lograr.

En el presente trabajo se aplica la metodología de la figura 1.12, la cual incluye el preprocesamiento, procesamiento y análisis de datos. A continuación, se detalla los métodos de preprocesamiento espectral utilizados.

1.7.1. Normalización

La normalización del hipercubo corrige la diferencia de escala en los valores de los píxeles. Diferentes métodos de normalización tiene cierto impacto en la clasificación (Cao et al., 2017), algunos de ellos son: normalización máxima, máxima-mínima, de banda máxima y de banda máxima-mínima (Cao et al., 2017). En la ecuación 2, la normalización máxima toma cada valor de los píxeles de una banda x_{ij} y lo divide por el valor máximo ($\max(x_{ij})$) del píxel de dicha banda.

$$x_{ij}^* = (x_{ij})/\max(x_{ij}) \quad (2)$$

En la normalización máxima-mínima que se muestra en la ecuación 3 se tiene el valor mínimo de la banda ($\min(x_{ij})$) el cual resta al píxel x_{ij} , esta diferencia se divide entre la resta $\max(x_{ij})$ y $\min(x_{ij})$ para cada banda.

$$x_{ij}^* = (x_{ij} - \min(x_{ij})) / (\max(x_{ij}) - \min(x_{ij})) \quad (3)$$

La normalización de una banda máxima se observa en la ecuación 4. En ella se elige el valor máximo de toda las bandas de los píxeles $\text{bandmax}(x_{ij})$ y lo divide a cada valor de píxel x_{ij} por banda.

$$x_{ij}^* = x_{ij} / \text{bandmax}(x_{ij}) \quad (4)$$

Por otra parte la ecuación 5 describe una normalización entre bandas máxima y mínima, donde la diferencia de ($\text{bandmax}(x_{ij}) - \text{bandmin}(x_{ij})$) actua como denominador frente a ($x_{ij} - \text{bandmin}(x_{ij})$).

$$x_{ij}^* = (x_{ij} - \text{bandmin}(x_{ij})) / (\text{bandmax}(x_{ij}) - \text{bandmin}(x_{ij})) \quad (5)$$

1.7.2. Filtro suavizante Savitzky-Golay

El método de suavizado Savitzky-Golay (Sv-G) es utilizado para mejorar la relación señal/ruido (Savitzky and Golay, 1964). El método Savitzky-Golay indica que para un conjunto de n puntos (x_j, y_j) , $y_j = f(x_j)$ en un plano se puede realizar un suavizado con la ecuación 6, donde $f(x_j)$ es el valor original y $g(x_j)$ el valor suavizado de combinación lineal de c_i y $f(x_{j+i})$; c_i son valores que depende del grado de polinomio que deseamos ajustar. Para un ajuste de “ m ” puntos se seleccionan una ventana móvil o región donde se encontrarán los “ m ” puntos en el rango de $-(m-1)/2 < i < (m-1)/2$. Se le llama ventana móvil debido que para los puntos que toma $(m-1)/2 \leq j < n - (m-1)/2$, se desplaza un punto hacia la derecha.

$$g(x_j) = \sum_{i=-(m-1)/2}^{(m-1)/2} C_i f_{x_{j+i}} \quad , \quad \frac{m-1}{2} \leq j \leq n - \frac{m-1}{2} \quad (6)$$

Permitiendo que se pueda reducir el ruido en los espectros, como se observa en la figura 1.13 que suaviza un espectro de días del año Vs índice de vegetación de diferencia normalizada (NDVI por las siglas en inglés *Normalized Difference Vegetation Index*) de una colección de imágenes.

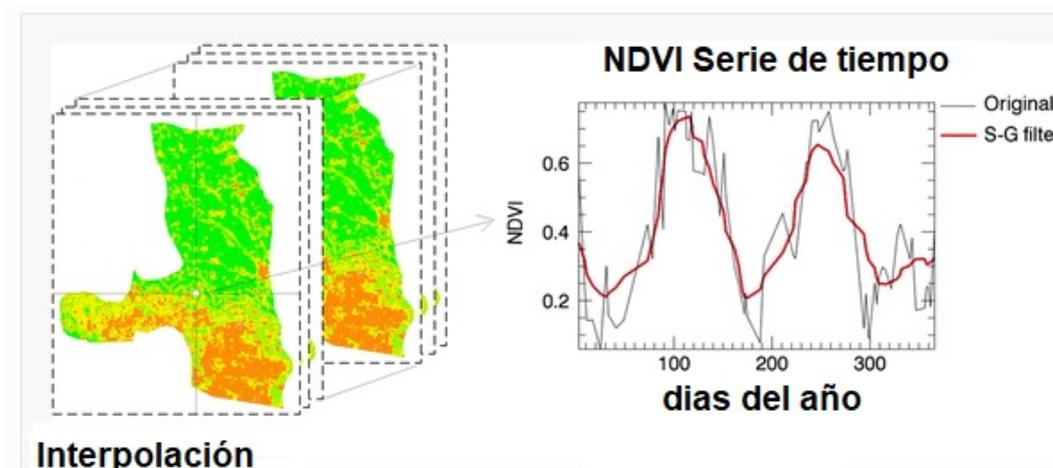


Figura 1.13: Corrección Savitzky para un espectro de NDVI.

1.7.3. Derivadas en el espectro de reflectancia

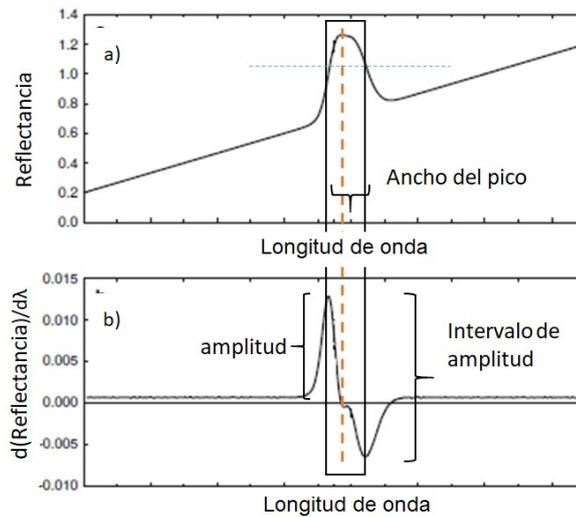


Figura 1.14: a) Espectro de reflectancia, b) derivada del espectro de reflectancia.

Las derivadas espectrales muestran la pendiente de una señal en un punto determinado y se aplican para eliminar los efectos de desviación vertical y de efecto de pendiente, así como para eliminar la línea base y extraer características espectrales como la forma espectral y el ancho de pico (Park and Lu, 2015). En la figura 1.14 se tiene el espectro original, el cual tiene un máximo valor de reflectancia, cuando se realiza la primera derivada al espectro, el valor máximo de reflectancia es igual a cero. Otra característica importante que se puede observar por diferenciación es que el ancho de un pico del espectro de referencia es inversamente proporcional a la amplitud del espectro derivado. Lo antes mencionado se puede ver en la figura 1.14: Cuando una señal tiene un pico ancho, la amplitud de su derivada en la ubicación del pico es pequeña (Fan et al., 2019). Por tal motivo la primera y segunda derivada son útiles para cuantificar los cambios de un espectro. En algunos casos es más útil utilizar la segunda derivada, si es que la señal no es afectada por el ruido.

1.7.4. Corrección de línea base

Un cambio de línea de base (o desplazamiento) se refiere a una señal de fondo de baja frecuencia que varía lentamente. Un cambio de línea de base distinto de cero introducido en los espectros medidos es causado por muchas fuentes diferentes; como el instrumento, la dispersión de la luz y diferentes longitudes de trayectoria espectral a través de la muestra. El objetivo de la corrección de línea de base es volver a colocar la línea de base en la línea de base de absorbancia cero (o reflectancia) restando la línea de base modificada de los datos espectrales medidos. La corrección de línea de base se

puede integrarse como parte de la normalización de los datos. En la figura 1.15 se tiene el espectro original con su línea de tendencia (de-trend), esta línea base es diferente de cero. Cuando se corrige la línea base, la pendiente es cero. En algunos casos eliminar la tendencia de los datos depende de las muestras que se analizan (MathWorks, 2019).

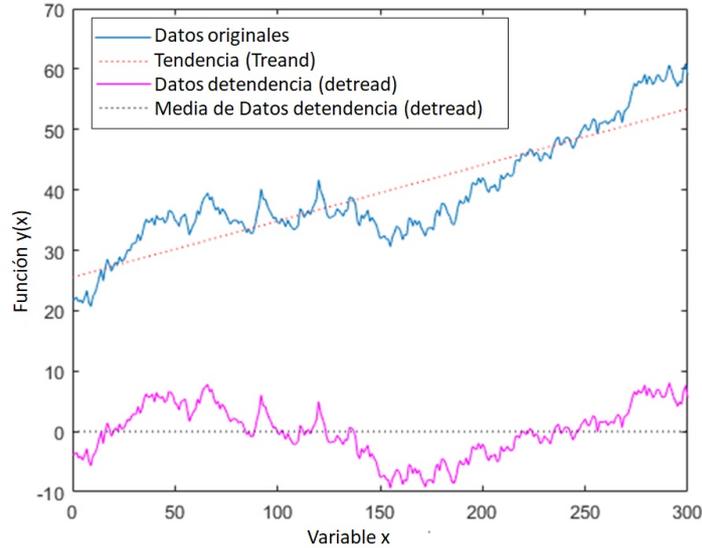


Figura 1.15: Cambio de línea base por de-trend (MathWorks, 2019).

1.7.5. Corrección de escala multiplicativa

La dispersión de la luz es una de las principales causas en la diferencia de escala multiplicativa. Por esta razón se utiliza la variable normal estándar (SNV); que reduce los efectos de dispersión. (Cao et al., 2017). La transformación SNV no necesita un espectro de referencia porque se aplica a muestras individuales. Para la transformación SNV, el espectro de cada muestra está centrado en la media y escalado por los valores derivados del propio espectro de la muestra, esto se observa en la ecuación 7.

$$\tilde{x}_{SNV(i)} = \frac{x_i - m(x)}{\sigma_i} \quad (7)$$

Donde m y σ son los valores promedio y de desviación estándar del espectro de muestra x_i , respectivamente y $\tilde{x}_{SNV(i)}$ es el SNV calculado de los datos sin procesar. El espectro corregido de SNV siempre tiene media cero y varianza unitaria. El cálculo de SNV corrige los cambios en el eje y . Sin embargo, los valores de SNV aún pueden ser afectados por la línea base, por lo tanto es necesario utilizar un polinomio de segundo orden (ver ecuación 8).

$$\hat{x}_{SNV(i)} = A_{SNV} + B_{SNV}W + C_{SNV}W^2 \quad (8)$$

Los valores para la regresión son los valores de reflectancia para el eje y y las

longitudes de onda para el eje \mathbf{x} . \mathbf{A} , \mathbf{B} y \mathbf{C} son los coeficientes de la regresión, \mathbf{W} la variable y $\hat{x}_{SNV(i)}$ el valor del ajuste.

Los valores ajustados por SNV y de-trend se muestra en la ecuación 9.

$$\tilde{x}_{SDT(i)} = \tilde{x}_{SNV(i)} + \hat{x}_{SNV(i)} \quad (9)$$

Donde $\tilde{x}_{SDT(i)}$ es el SNV-D. Con el SNV-D removemos la linea base originados por el NIR (Barnes et al., 1989).

1.8. Procesamiento

El procesamiento de imágenes hiperespectrales se refiere al uso de algoritmos informáticos para extraer, almacenar y manipular información de las imágenes hiperespectrales para el procesamiento de información, análisis de datos, clasificación de datos, regresión, detección de objetivos y reconocimiento de patrones. Debido que las imágenes hiperespectrales contienen información espectral y espacial, se han desarrollado herramientas y técnicas especiales para procesar ambas.(Manolakis et al., 2016; Amigo et al., 2015; Park and Lu, 2015).

Desde la perspectiva de la estadística multivariada, las imágenes hiperespectrales son datos multivariados; por lo tanto las técnicas de análisis multivariadas se aplican para procesar las imágenes hiperespectrales como el análisis de componentes principales y el análisis de regresión de mínimos cuadrados parciales; otra manera de procesar imágenes hiperespectrales es utilizando aprendizaje por máquina y algoritmos para abordar problemas más complejos, como la detección en tiempo real de los materiales. A continuación, se menciona algunos preprocesamientos que son mayormente utilizados.

1.8.1. Análisis de componentes principales (PCA)

La técnica de análisis de componentes principales (PCA) es una técnica estadística en el que se utiliza una matriz de datos el cual contiene una cantidad de individuos u observaciones a los que se mide múltiples variables. En el caso de las imágenes hiperespectrales, la matriz aleatoria es la matriz de datos y ésta se representa de la siguiente manera:

$$X = (x_{ij}) = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1p} \\ \cdot & & & & \cdot \\ x_{i1} & \dots & x_{ij} & \dots & x_{ip} \\ \cdot & & & & \cdot \\ x_{n1} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix} \quad (10)$$

Donde x_{ij} son los píxeles. El PCA se utiliza para la reducción de dimensiones o variables debido a que algunas de ellas se pueden correlacionar; por lo tanto, es necesario reducirlo. El PCA se basa en que las p -variables (en las HSI las p -variables son las bandas) pueden tener información en común o estar correlacionadas, se busca que estas p -variables se puedan transformar en otras nuevas variables no correlacionadas entre sí llamadas componentes principales.

Estas nuevas variables no correlacionadas son combinaciones lineales de las anteriores variables y están ordenadas de máxima variabilidad a mínima variabilidad.

■ Cálculo de las componentes principales

A partir de la matriz aleatoria se construye el vector aleatorio p -dimensional o variables de la forma siguiente $X = [X_1, X_2, \dots, X_p]^T$, a partir de estas variables se calcula un nuevo conjunto de variables Y_1, Y_2, \dots, Y_p no correlacionadas entre sí y con varianzas que decrecen progresivamente.

Cada una de estas nuevas variables Y_j ($j=1,2,\dots,p$) son las llamadas componentes principales y son combinaciones lineales de las componentes originales X_j como se observa en la ecuación 11.

$$Y_j = a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jp}X_p = a'_j X \quad (11)$$

Siendo $a'_j = (a_j, a_{2j}, \dots, a_{pj})$ un vector constante y que cumple la siguiente condición para cada nueva componente (en el caso de las imágenes se tiene 168 componentes):

$$a'_j a_j = \sum_{k=1}^p a_{kj}^2 = 1 \quad (12)$$

$a_j = (a_j, a_{2j}, \dots, a_{pj})^T$ La condición que se debe cumplir para que la nueva variable sea la primera componentes es: $\text{Var}(Y_j)$ sea máxima y así sucesivamente.

Como ejemplo se pueden tomar cuatro píxeles de dos bandas que corresponderían a las variables X_1 y X_2 con sus respectivos valores x_k que se ubican en la figura 1.16, en la cual se puede trazar infinitas rectas Y , siendo y_{1k} la recta con mayor variabilidad al proyectar los puntos x_k en esta recta, en los puntos P_k , donde a_1 es el vector que contiene las constantes para la recta y_{1k} . Se tiene como ejemplo la figura 1.16 de dos componentes, los valores de estas componentes se proyectan en una de las infinitas rectas que se pueden trazar con vector constante a_1 .

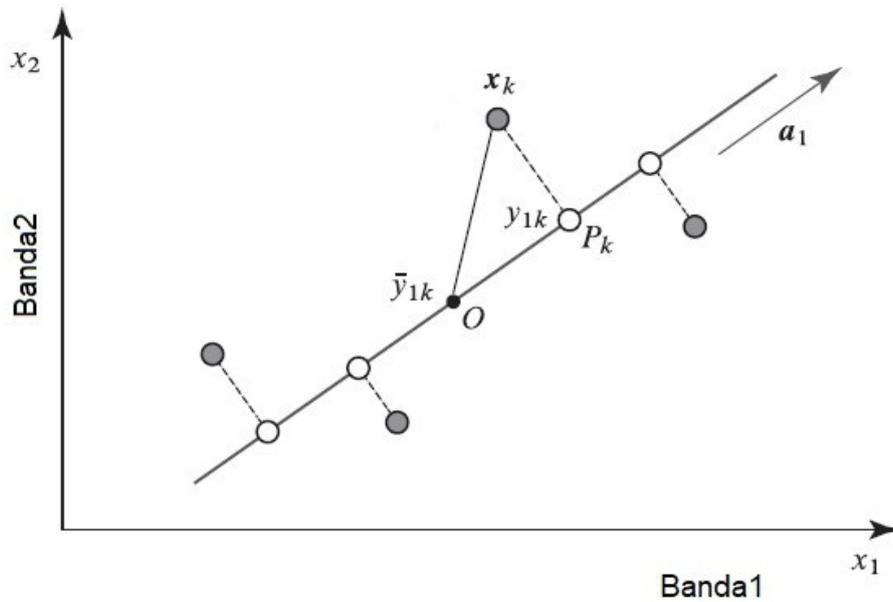


Figura 1.16: PCA como una proyección ortogonal en una dimensión inferior: puntos proyectados en una línea (Manolakis et al., 2003).

1.8.2. Prueba de hipótesis para clasificar imágenes hiperespectrales

En la figura 1.17 se tiene dos funciones normales obtenidas a partir de dos imágenes: una de suelo contaminado y otra de suelo no contaminado. En esta figura representa la prueba de hipótesis para clasificar imágenes hiperespectrales. En estas funciones se pueden identificar dos puntos (máximos) con la mayor densidad (probabilidad) que ubican los valores medios, las modas y las medianas de los suelos en el eje horizontal. A estos puntos se les denomina respectivamente hipótesis H_0 llamada hipótesis nula o contaminante ausente e hipótesis alternativa H_1 o contaminante presente. (Manolakis et al., 2016; Park and Lu, 2015). La distancia entre H_0 y H_1 (D) aumenta a mayor concentración del suelo contaminado. Las áreas de color negro y rojo son conocidas como error de tipo 1 y tipo 2, respectivamente. Estas áreas se interpretan como : las cantidades de píxeles que no se pueden clasificar como no contaminado y contaminado para el área negra y roja, respectivamente. El nivel crítico o límite separa estas dos áreas, cuando el nivel crítico es igual a cero, todo los píxeles se pueden clasificar.

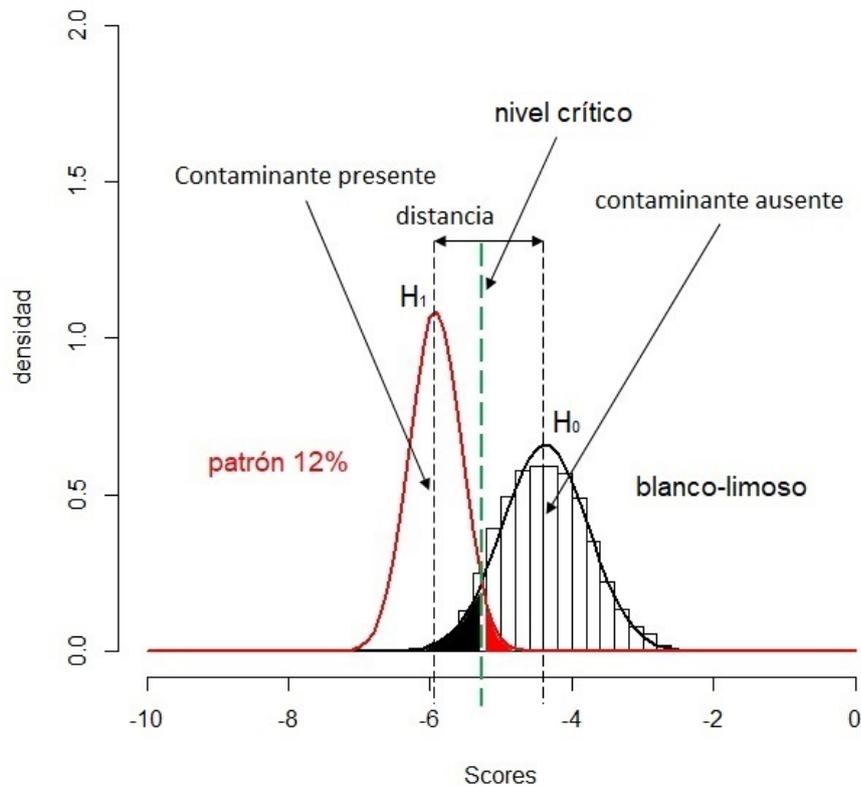


Figura 1.17: Prueba hipótesis entre suelo limoso y el suelo limoso con crudo de petróleo a 12%pp. Se muestran las funciones de densidad y la intersección entre ellas.

El desplazamiento de la función de suelo contaminado frente a la función de suelo no contaminado permite analizar la distancia entre las hipótesis y el nivel crítico. El área de color negro se refiere al porcentaje de suelo no contaminado que no se puede clasificar. El área de color rojo se refiere al porcentaje del suelo contaminado que no se puede clasificar.

1.9. Conceptos generales de espectrometría en el infrarrojo

Las regiones que conforman el espectro electromagnético se dividen en ondas de radio, microondas, infrarrojo, luz visible, ultravioleta, rayos x y rayos γ , todas estas regiones son parte del espectro electromagnético. Como se muestra en la (figura 1.18) las ondas de radio son las que tienen una mayor longitud de onda (λ) y los rayos gamma presentan las longitudes de onda más cortas de todo el espectro electromagnético (Hecht, 2000; Serway and Jewett, 2018).

Con respecto a los límites no existe una división precisa entre las regiones que forman el espectro electromagnético; pero se puede dividir en diferentes regiones cuyos

límites son más o menos arbitrarios. En la tabla 1.2 se indican los rangos aproximados de frecuencias (en Hz), longitudes de ondas (en metros) y energías (eV) de las diferentes regiones del espectro electromagnético.

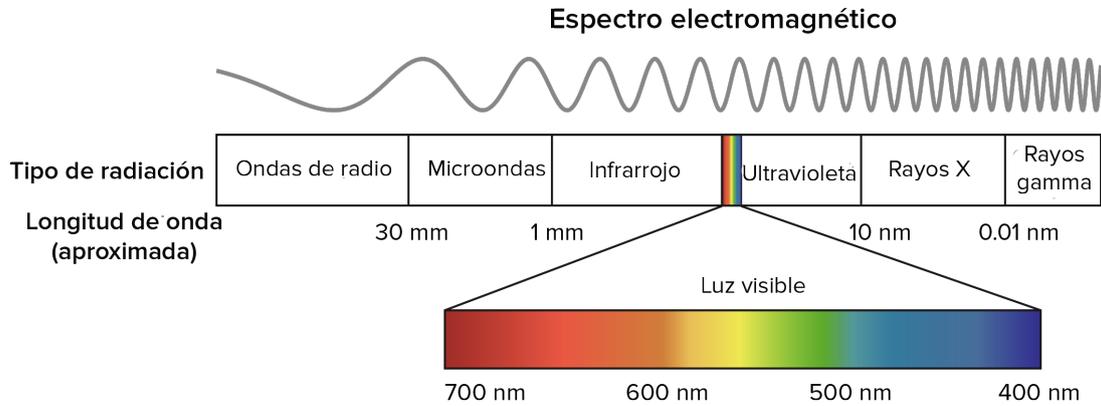


Figura 1.18: Espectro electromagnético. Adaptado de Hecht (2000).

Cada una de estas regiones son utilizadas en diferentes mediciones para encontrar una respuesta asociada a un valor físico, químico o biológico que se desea medir; una de las regiones que capta mayor atención es la región NIR, la que fue la primera zona del espectro de absorción no visible que se descubrió y la cual se encuentra entre las regiones de microonda y visible. La región infrarrojo se divide en tres zonas debido a razones instrumentales y características de la interacción de la radiación con la materia, las regiones que se divide son : infrarrojo próximo (NIR: *Near infrared*), infrarrojo medio (MIR: *Middle infrared*) e infrarrojo lejano (FIR: *Far infrared*). Los orígenes de absorción en el infrarrojo se da en la parte molecular debido a sobretonos y bandas de combinación de vibración moleculares fundamentales para el NIR, vibraciones moleculares fundamentales para el MIR y rotaciones moleculares para el FIR.

Tabla 1.2: Regiones del espectro electromagnético con sus respectivos intervalos de frecuencia, longitud de onda y energía (Hecht, 2000)

Región	Frecuencia (Hz)	Longitud de onda (m)	Energía de fotón (eV)
Rayos gamma	10^{22} - 10^{19}	10^{-11} - 10^{-14}	10^{10}
Rayos X	10^{17} - 10^{20}	10^{-8} - 10^{-11}	10^5
Ultravioleta	10^{15} - 10^{17}	10^{-6} - 10^{-8}	10^3
visible	10^{14} - 10^{15}	4×10^{-6} - 7×10^{-6}	1
Infrarrojo	10^{11} - 10^{14}	10^{-3} - 10^{-5}	10^{-3}
Microonda	10^9 - 10^{11}	10^{-1} - 10^{-3}	10^{-5}
Ondas de Radio	10^3 - 10^9	10^5 - 10^{-1}	10^{-10}

1.9.1. Medidas en infrarrojo

Las medidas que se observan en el infrarrojo son la absorción y la reflectancia (reflectancia difusa o reflectancia especular). La reflectancia especular ocurre cuando la luz incidente en la superficie muy plana se refleja formando un ángulo incidente igual al ángulo reflejado. La reflectancia difusa ocurre cuando la luz incidente cae en la superficie rugosa y refleja la luz con ángulo disperso. De forma microscópica en la reflectancia difusa un rayo incidente se refleja con ángulo idéntico al incidente. Debido a la naturaleza dual onda-partícula de las radiaciones electromagnéticas, se han realizado modelos teóricos para describir la respuesta espectral infrarroja. A continuación, se muestra el modelo de luz sin dispersión.

■ Modelo de luz sin dispersión

Este modelo es usado para calcular la concentración de un soluto en gases y líquidos. Fue desarrollado por Pierre Bouguer (1729) y Johann Lambert (1760) de manera independiente y es conocido como la ley de Lambert-Beer (Park and Lu, 2015), donde la intensidad I_0 que atraviesa una longitud l disminuye en forma exponencial en relación al coeficiente de absorción (k), la concentración del componente ($[s]$) y la longitud que atraviesa el haz de luz (ecuación 13).

$$\frac{I_l}{I_0} = e^{-k[s]l} \quad (13)$$

Tomando el logaritmo en base e a cada lado de la ecuación 13, se tiene la siguiente ecuación 14.

$$\begin{aligned} \ln\left(\frac{I_0}{I_l}\right) &= k[s]l \\ \ln\left(\frac{1}{T}\right) &= A \end{aligned} \quad (14)$$

La ecuación 14 relaciona la transmitancia ($T = \frac{I_l}{I_0}$) con la absorbancia ($A = k[s]l$). La relación que existe entre la concentración $[s]$ y la absorbancia (A) está descrita por la *ley de Lambert-Beer*.

En el análisis de reflectancia difusa se toma la ley de Lambert-Beer y se sustituye la energía transmitida (T) por la energía reflejada (R). La nueva ecuación obtenida $\ln\left(\frac{1}{R}\right) = A$ permite relacionar la reflectancia con la absorbancia.

CAPÍTULO 2

2. Suelos e hidrocarburos de petróleo

2.1. Suelos y propiedades espectrales del suelo

El suelo es la capa más superficial de la corteza terrestre. La composición del suelo es una mezcla compleja de minerales, materia orgánica, gases y agua en diferentes proporciones.

2.1.1. Componentes básicos del suelo

En general los suelos tienen componentes sólidos (minerales y materia orgánica), líquidos y gaseosos:

- **Fase sólida.**-Esta fase está dividida en orgánica e inorgánica. La parte orgánica está compuesta por madera, plantas, entre otros. La fase sólida inorgánica está compuesta por minerales y rocas fragmentadas.
- **Fase líquida.**- En la fase líquida se encuentran las sales minerales, coloides de arcillas y humus. El líquido se encuentra en los poros del suelo de tamaño pequeño o mediano.
- **Fase gaseosa.**- Es el espacio que ocupa los poros de tamaño grande y donde el líquido se ha consumido. Se compone principalmente de O_2 y CO_2 .

2.1.2. Propiedades físicas o mecánicas del suelo

Las propiedades físicas o mecánicas del suelo están relacionadas con la proporción de los diferentes componentes que la conforman. Los suelos presentan las siguientes propiedades físicas: textura, estructura, consistencia, densidad, aireación, temperatura, color y porosidad .

- **Textura del suelo**

La textura de un suelo se define por la proporción de arena, limo y arcilla y por el tamaño de las partículas. Para clasificar el suelo según su textura, se utiliza el triángulo de textura (ver figura 2.1) diseñado por el departamento de agricultura de los Estados Unidos de América (USDA) (Gisbert Blanquer et al., 2010). En los extremos de triángulo de textura se encuentran los suelos puros (limoso, arcilloso y arenoso). La combinación de los diferentes suelos puros en diferentes

proporciones completan el triángulo de textura con los suelos arcillo arenoso, franco arcillo arenoso, franco arenoso, areno franco, franco, franco arcilloso, arcillo limoso, franco arcillo limoso y franco limoso.

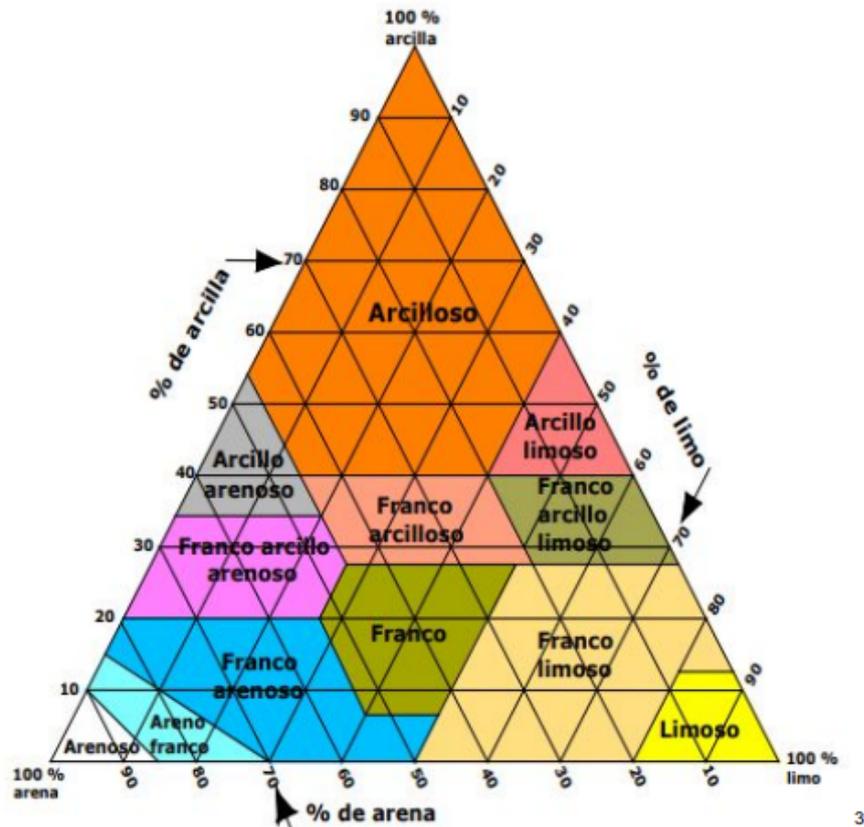


Figura 2.1: triángulo de textura de suelo diseñado por USDA (Gisbert Blanquer et al., 2010).

- **Estructura del suelo**

Esta propiedad se define en base a los agregados de los suelos. Los agregados son la agrupación de partículas individuales, tomando el aspecto de partículas más grandes. Se distinguen suelos de estructura esferoidal (agregados redondeados), laminar (agregados en láminas), prismática (en forma de prisma), bloques y granular (en granos).

- **Consistencia del suelo**

Se refiere a la resistencia para la deformación o ruptura. Según la resistencia el suelo puede ser suelto, suave, duro y muy duro.

- **Densidad del suelo**

Se refiere al peso por volumen del suelo y está relacionado con la porosidad. Un suelo muy poroso será menos denso y un suelo poco poroso será más denso. A

mayor contenido de materia orgánica es más poroso y menos denso será el suelo. Mediante la determinación de la densidad se puede obtener la porosidad total del suelo.

■ **Color**

El color del suelo depende de sus componentes y puede usarse como una medida indirecta de ciertas propiedades como la humedad. Por ejemplo: el color rojo de un suelo indica el contenido de óxidos de hierro y manganeso; el amarillo indica óxidos de hierro hidratado; el blanco y el gris indican presencia de cuarzo, yeso o caolín; el suelo negro y marrón indican materia orgánica.

■ **Porosidad del suelo**

El espacio poroso del suelo se refiere al porcentaje del volumen del suelo no ocupado. En general el volumen del suelo está constituido por 50 % de material sólido (45 % minerales y 5 % materia orgánica) y 50 % de espacios porosos. Dentro de los espacios porosos se pueden distinguir macro poros y micro poros donde el agua, los nutrientes, el aire y los gases pueden circular o retenerse.

■ **Temperatura**

La temperatura del suelo es importante porque determina la distribución de las plantas e influye en los procesos bióticos y químicos.

■ **Aireación**

Se refiere al contenido de aire del suelo y es importante para el abastecimiento de oxígeno, nitrógeno y dióxido de carbono en el suelo.

2.1.3. Propiedades espectrales de los suelos

Para analizar los espectros de los suelos, se tiene que considerar la respuesta espectral de cada componente del suelo (Winkelmann, 2005).

La presencia del agua en un suelo se manifiesta por la presencia de dos bandas de absorción en 1.4 μm y 1.9 μm . En los suelos la reflectancia es más intensa hacia los extremos de los espectros entre las longitudes de onda de 0.4 μm y 2.5 μm . Las características de los materiales orgánicos en el suelo se pueden encontrar entre las longitudes de onda de 0.5 μm - 2.0 μm y la presencia de minerales se manifiesta principalmente entre las longitudes de onda de 2.1 μm y 2.5 μm (Winkelmann, 2005; Manolakis et al., 2003).

En la figura 2.2 se muestra espectros de algunos suelos: (E) Suelo orgánico (contiene carbón orgánico \gg 4 %, óxido de hierro $<$ 1 %), (B) suelo orgánico mínimamente alterado (contiene carbón orgánico $<$ 2 %, óxido de hierro $<$ 1 %), (C) suelo afectado

por hierro (contiene carbón orgánico < 2 %, óxido de hierro 1,4 %), (A) suelo orgánico afectado (contiene carbón orgánico > 2 %, óxido de hierro < 1 %) y (D) Suelo con dominio de hierro (contiene óxido de hierro >> 4 %) (Winkelmann, 2005).

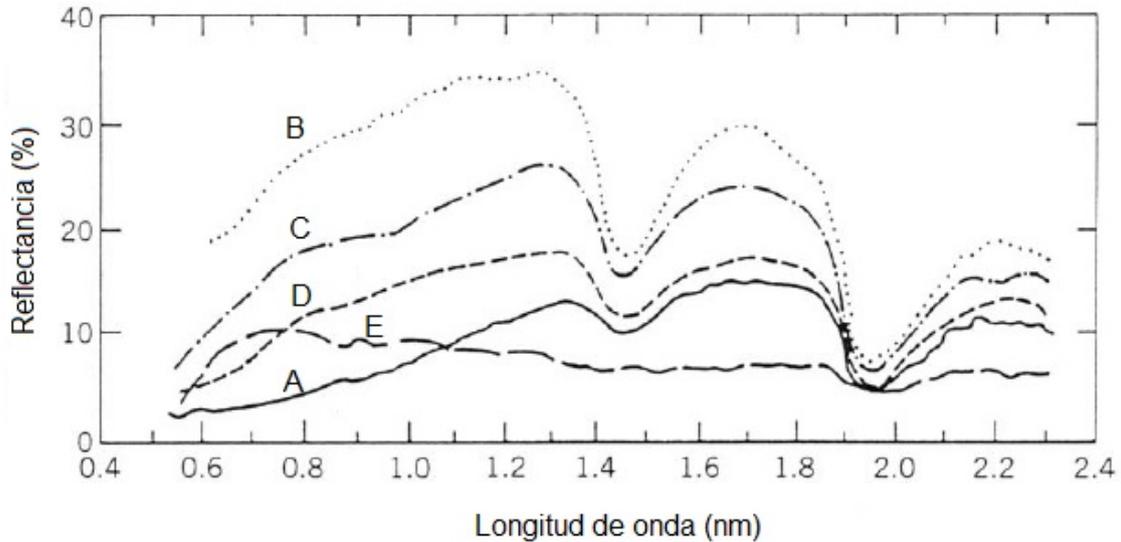


Figura 2.2: Reflectancia típica de cinco tipos de suelo: (E) orgánico, (B) orgánico mínimamente alterado, (C) afectado por hierro, (A) orgánico afectado y (D) con dominio de hierro. Adaptado de Winkelmann (2005).

2.2. Hidrocarburos de petróleo (PHC)

Bajo el término de hidrocarburos de petróleo (PHC por las siglas en inglés *Petroleum Hydrocarbon*) se engloba cualquier sustancia presente en petróleo o derivados del mismo. Los hidrocarburos de petróleo corresponden a una amplia serie de compuestos que se dividen en tres grandes grupos: los alcanos (parafinas), los alquenos y los aromáticos. La característica de los alcanos son que carecen de enlaces dobles o triples. Los alquenos tienen como característica dos enlaces carbono-carbono en su molécula. Los aromáticos son compuestos orgánicos cíclicos que posee una mayor estabilidad molecular (Botello, 2005).

2.2.1. Origen y propiedades

El petróleo crudo tiene una composición química muy compleja que varía según su origen formados por sustancias hidrocarbonadas que constituyen del 50 al 90 % del total y son del tipo n-alcanos, alcanos ramificados, cicloalcanos e hidrocarburos aromáticos. Otros elementos a menor cantidad son N, S y O como también algunos metales.

Dentro de los hidrocarburos más comunes se pueden mencionar: la gasolina, el diésel, el crudo y aceites lubricantes. El crudo está compuesto entre 50 y 98 % de hidrocarburos

y puede contener también otros elementos (Botello, 2005).

2.2.2. Contaminación del suelo por hidrocarburos de petróleo

La contaminación del suelo se debe a la presencia de sustancias que dada una concentración degrada el suelo produciendo una pérdida total o parcial de las propiedades físicas, químicas y biológicas. Los tipos de contaminación en el suelo se clasifican como: Contaminación natural o endógena y contaminación antrópica o exógena. (Felipó, 1992)

- **Contaminación natural o endógena**

En este tipo de contaminación se observa el proceso de concentración y toxicidad que muestran los elementos metálicos.

- **Contaminación antrópica o exógena**

Se refiere a una contaminación que es introducida al medio (atmósfera, suelo o agua) debido a la actividad del hombre.

Para el estudio de suelos contaminados se tiene que analizar los posibles factores que influyen en la respuesta del suelo a los contaminantes, los cuales son:

- **Vulnerabilidad:** Se refiere a la sensibilidad del suelo frente al contaminante. El grado de vulnerabilidad depende de la intensidad de afección, el tiempo y la velocidad con la que se producen los cambios del suelo.

- **Poder de amortiguación:** El suelo puede actuar como barrera protectora de otros medios (biológicos e hidrológicos) y pueden presentar una elevada capacidad de depuración o amortiguamiento de contaminantes.

Cuando el suelo alcanza el límite de su capacidad de depuración, deja de ser eficaz, llegando a ser un suelo contaminado.

2.2.3. Propiedades espectrales de los hidrocarburos de petróleo

Las frecuencias propias de algunas frecuencias del NIR pueden coincidir con las frecuencias naturales de los enlaces que conforman los hidrocarburos (ejemplo C-H). Al coincidir ambas frecuencias, la molécula pasa de estar en un estado de equilibrio E_0 a un estado superior E_i , $i=1,2,\dots$ (ver figura 2.3). La distancia entre partículas r_0 cambia para un estado superior E_i . La absorción en una molécula de hidrocarburo se da cuando ocurre el cambio de estado. Esta absorción se puede observar en un espectro de reflectancia (ver 2.4).

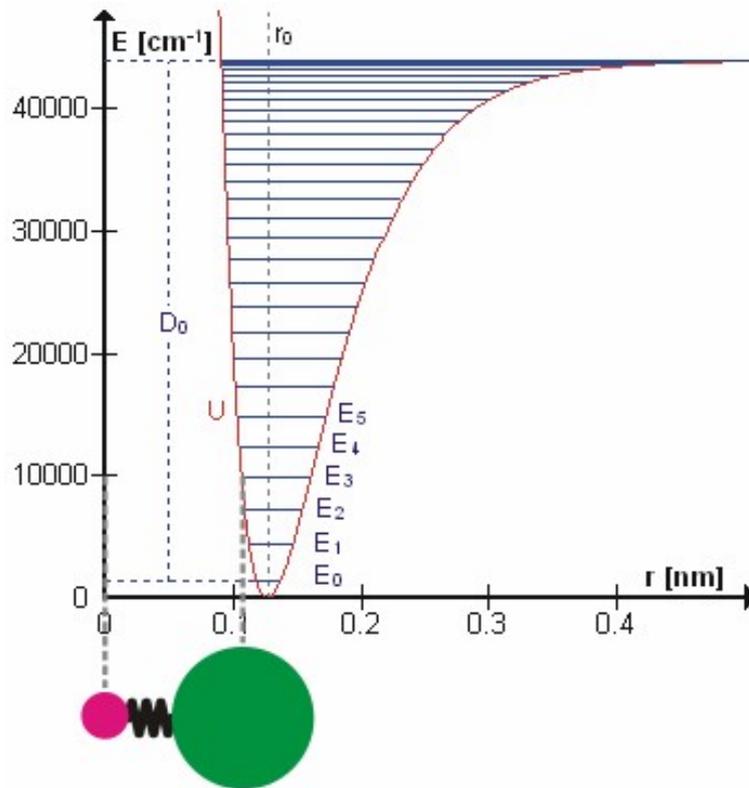


Figura 2.3: La radiación infrarroja absorbida por una molécula causa que los enlaces individuales vibren en gran aproximación como osciladores diatómicos

El enlace que caracteriza a los hidrocarburos es el enlace carbono-hidrógeno (C-H) y presentan absorción alrededor de las longitudes de onda de $2.3 \mu\text{m}$, $1.7 \mu\text{m}$, y $1.2 \mu\text{m}$ (ver figura 2.5).

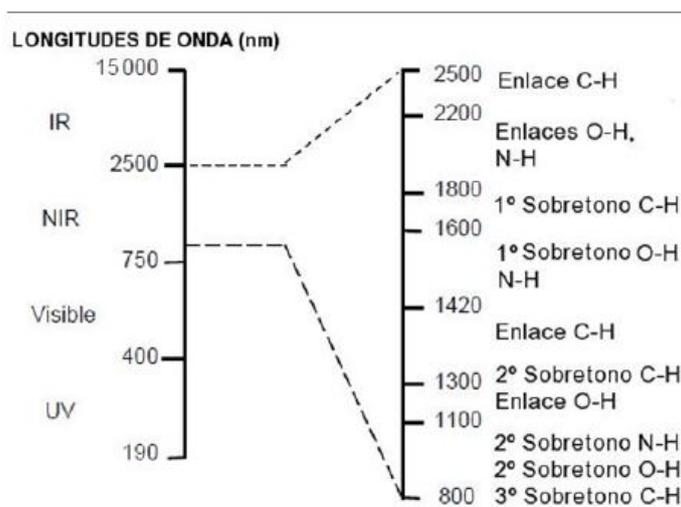


Figura 2.4: Bandas de absorción en el NIR (Martínez, 2015).

Los enlaces del tipo $C - H$, $O - H$, $N - H$ poseen longitudes de onda de absorción características en la región espectral entre 780 a 2500 nm (Siesler et al., 2008). En la figura 2.4 se muestra la región NIR entre 750 a 2500 nm y los enlaces C-H, O-H y N-H asociados que se dan en ciertas longitudes de onda.

Para el propósito de la detección de imágenes hiperespectrales en la investigación de hidrocarburos o derivados de petróleo, es necesario considerar la distribución espacial del hidrocarburo sobre la superficie del suelo.

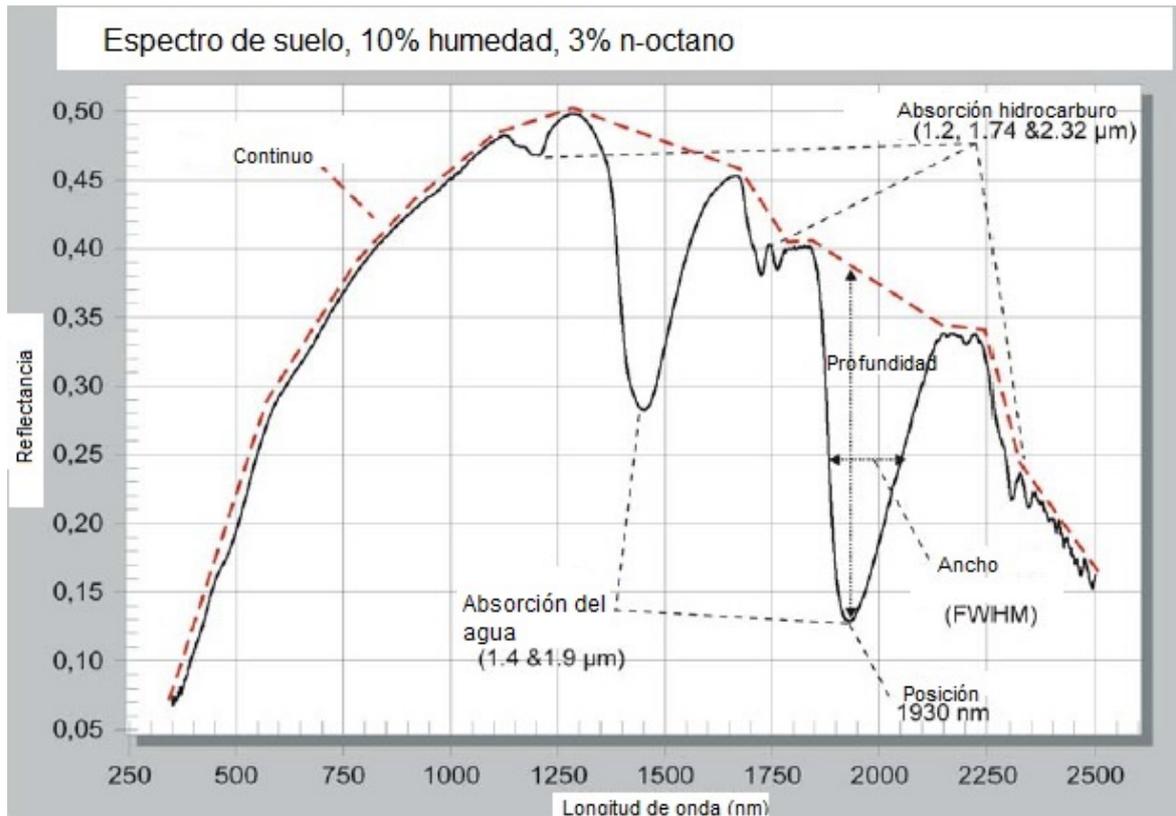


Figura 2.5: Características de un espectro en el infrarrojo cercano (NIR) de un suelo contaminado con hidrocarburo. Adaptado de Winkelmann (2005).

En la figura 2.5 se observa el espectro del suelo contaminado con 10 % de humedad y 3 % de hidrocarburo. En el espectro, las longitudes de onda que caracterizan el agua son: 1.4 μm y 1.9 μm . Las longitudes que caracterizan a los hidrocarburos son 1.2 μm , 1.74 μm y 2.32 μm .

CAPÍTULO 3

3. PROCEDIMIENTOS Y RESULTADOS

Los resultados obtenidos se dividen en cinco partes: Calibración del sistema de imágenes hiperespectrales, preparación de los patrones de referencia (PR), adquisición de imágenes hiperespectrales, preprocesamientos de las imágenes hiperespectrales y procesamientos de las imágenes hiperespectrales. La metodología de trabajo se muestra en la figura 1.12. Para cada etapa de la metodología se utilizan diferentes métodos de análisis multivariado, software especializados como Spectronon-Pro y el lenguaje de programación *R* en la plataforma de R-Studio. Ver ANEXO A para el manejo del software Spectronon-Pro y ANEXO B para el código en lenguaje *R*.

3.1. Calibración del sistema de imágenes hiperespectrales (HSI)

La calibración del sistema HSI se realiza para garantizar la reproducibilidad de los valores de las imágenes hiperespectrales. La calibración del sistema HSI consiste en: (1) conocer la reflectancia de teflón, (2) calibración de la reflectancia o también llamado la corrección de campo plano para la cámara hiperespectral y (3) calibración de los parámetros del sistema HSI. Se utilizó un teflón espectralmente plano y de superficie homogénea de 27 cm de largo, 13.3 cm de ancho y 1.1 cm de espesor.

3.1.1. Reflectancia del teflón



Figura 3.1: Equipos utilizados para conocer la intensidad de luz reflejada por el teflón.

La reflectancia del teflón fue calculada de los datos de intensidades que se obtuvieron usando un fotodetector THORLABS-DET10D, una lámpara de luz halógena y un osciloscopio Tektronix-DPO2012 C001176, los cuales se observan en la figura 3.1. Previo

a la medición de reflectancia, se lijó el teflón con un papel de lija de 100 g para tener una superficie lisa. (Resonon, 2017).

Cuando esté posicionado el fotodetector frente a la fuente de luz como se observa en la figura 3.2 A, el osciloscopio registra un valor proporcional a la intensidad de luz incidente I_0 en mV . De la misma manera se busca la intensidad reflejada I_R del teflón cuando la luz de la lámpara incide sobre él, esto se puede ver en la figura 3.2 B.

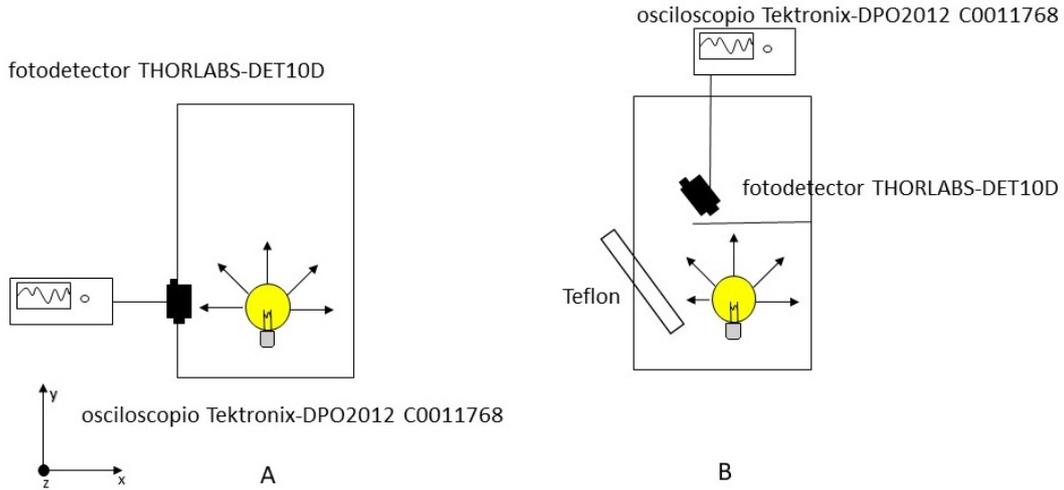


Figura 3.2: A. Esquema para medir la intensidad de luz incidente, B. Esquema para medir la intensidad de luz reflejada por el teflón.

Se realizaron 10 mediciones de I_0 e I_R . En la tabla 3.1 se muestra los valores medios de I_0 e I_R . Así mismo la dispersión es dada por $D = \text{Valor máximo} - \text{valor mínimo}$, para cada medición. También se muestra $T_{10} = D/\bar{x}$, donde \bar{x} es el valor medio. En esta tabla se observa que los valores medios que se obtuvieron para I_0 e I_R son: 33.45 mV y 16.26 mV con dispersión de 7 mV y 1 mV respectivamente. La reflectancia del teflón (R) es la relación entre $\frac{I_R}{I_0}$.

Tabla 3.1: Registros en el sensor (en mV) para I_0 e I_R

N	Registro para I_0 (mV)	Registro final I_R (mV)
valor medio \bar{x}	33.45	16.26
D	7	1
T_{10}	0.21	0.06

Para obtener valores absolutos de reflectancia se puede usar el valor de referencia 0.49 ± 0.02 , es decir que para una reflectancia media \bar{I} su valor absoluto $I_{absoluto}$ es

$I_{absoluto} = I * 0.49$, en la tabla 3.2 se puede observar los valores tomados para calcular dicha reflectancia.

Tabla 3.2: Valores calculado de intensidad y error

I_0	I_R	R	σ
33.45	16.29	0.49	0.02

3.1.2. Sistema de imagen hiperespectral de laboratorio

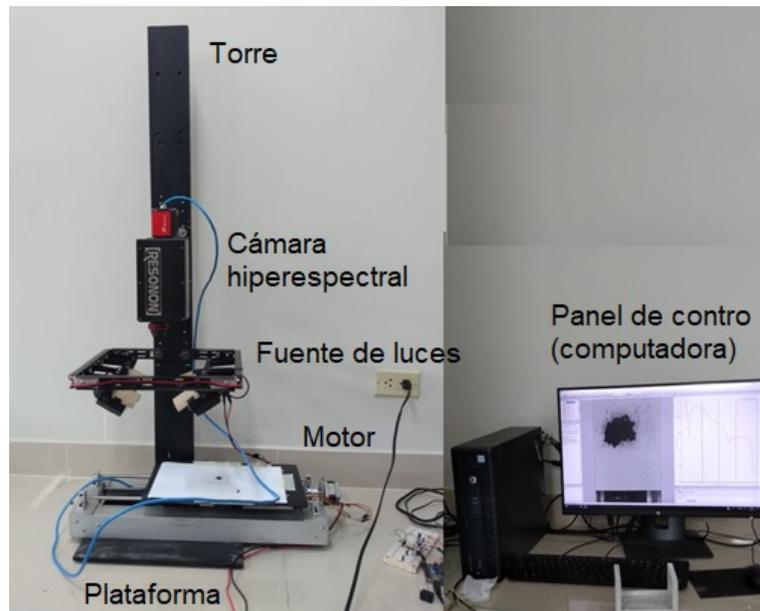


Figura 3.3: Sistema hiperespectral con cámara Pika NIR-320 de la marca RESONON.

El sistema óptico que se utilizó para registrar las imágenes consta de: una cámara hiperespectral Pika NIR-320 de la marca Resonon con un lente de longitud focal de 25 mm con un campo de visión (FOV) de 21.7° y un campo de visión instantánea (IFOV) de 1 mrad en la región espectral de 900 a 1700 nm, resolución espectral de 4.9 nm, 168 canales espectrales y otras características que se observan en la tabla 3.3.

Tabla 3.3: Características de la cámara Pika NIR-320

Rango espectral	900-1700 nm
Resolución espectral	4.9 nm
Canales espectrales	168
píxeles espectrales	164
Canales espaciales	320
Frecuencia de imagen máxima	520 fps
Profundidad de bits	14
Relación de señal a ruido (pico)	1885
Opciones de conexión	GigE
requerimientos de energía	12V, 5A

La cámara hiperspectral Pika NIR-320 se maneja con un sistema de laboratorio el cual se observa en la figura 3.3. Este sistema cuenta con una plataforma de movimiento en la cual se ubican las muestras, una torre que sostiene la cámara, las fuentes de luces halógenas y un motor que mueve la plataforma y es controlado por la computadora. Los cuales se describen a continuación.

- **Plataforma de muestra**

Su función es trasladar la muestra a través del campo de visión de la cámara. Cuenta con una plataforma de acrílico con dimensiones de 24 cm de ancho y 35.5 cm de largo y se mueve utilizando un motor a paso que es controlado por un Arduino-uno.

- **Torre del sistema**

Su función es servir como soporte del sistema de iluminación y de la cámara hiperspectral. Está hecha de aluminio y está asegurada en una base de gran peso para poder soportar los 6 kg de la cámara, la altura máxima que tiene es de 120 cm. La cámara se puede ajustar hasta esa altura y aumentar su resolución espacial.

- **Sistema de iluminación**

Su función es ser la fuente de radiación para la muestra. Se encuentra ubicada debajo del lente óptico de tal manera que no incida la luz directamente al lente. El sistema de iluminación cuenta con 4 unidades de lámparas halógenas de corriente directa con voltaje de 220 V y una potencia de 50 W.

- **Computadora**

El sistema de adquisición de imágenes hiperspectral cuenta con una computadora Intel(R) Core(TM) i7-6700 con una memoria RAM de 16 Gb y un disco duro

de 1 TB. La computadora controla la plataforma a través del circuito electrónico Arduino-uno y registra las imágenes por medio del software Spectronono-Pro. La computadora aparte de adquirir las imágenes también se usa para el preprocesamiento de una imagen hiperespectral, así como para su procesamiento, parte fundamental en el trabajo desarrollado.

3.1.3. Calibración del sistema hiperespectral

Para la adquisición de las imágenes, los parámetros que se utilizaron son: distancia del lente a la muestra de 39.3 cm, fotogramas por segundos de 52 Hz, velocidad de la plataforma de 2.3 cm/s, resolución espacial 0.48 mm, tiempo de integración 11.09 ms, dimensiones de las imágenes hiperespectrales de 500 x 320 píxel con 168 bandas y con un área de barrido de 15.2 x 22.9 cm^2 (ver tabla 3.4). La adquisición del HSI, se realiza en modo de detección de reflectancia NIR, y modo de escaneo de banda intercalada por línea (BIL). La parte de adquisición de imagen se realiza con el software *Spectronon-Pro*. El detalle del uso del software se muestra en el **ANEXO A**.

Tabla 3.4: Parametros del sistema hiperespectral

Distancia Lente-muestra (cm)	Framerate (Hz)	Velocidad plataforma (cm/s)	Resolución espacial (mm)	Tiempo de integración (ms)	Área barrido (cm ²)
39.3	52	2.3	0.48	11.09	15.2x22.9

La calibración de reflectancia se realiza mediante una función automática en el software *Spectronon-Pro*, esta corrección utiliza la intensidad del teflón o referencia $E_{referencia}$ y la intensidad de la corriente oscura o ruido de fondo $E_{corriente-oscura}$. El $E_{referencia}$ se adquiere del patrón de referencia de calibración (teflón), el software considera al blanco de referencia con un valor de reflectancia ~ 99.9 % y el $E_{corriente-oscura}$ se adquiere cubriendo el lente con una tapa oscura, reflectancia ~ 0.0 %, ver figura 3.4.

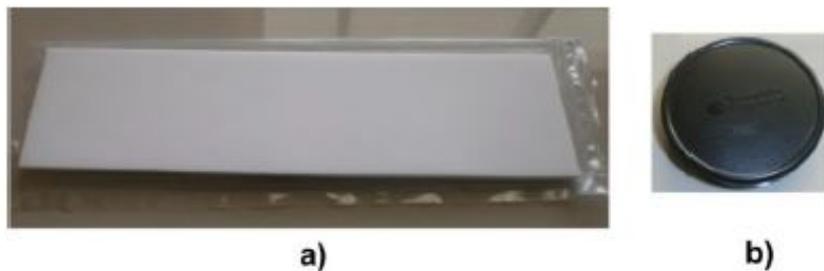


Figura 3.4: a) Patrón de teflón para $E_{referencia}$ b) Cubierta de lente para $E_{corriente-oscura}$.

La corrección de las imágenes hiperespectrales se realiza mediante la ecuación 15; donde I_{blanco} son las intensidades de corriente en el hipercubo del teflón, I_{oscuro} son las intensidades del hipercubo cuando se tapa el lente de la cámara e I_m el hipercubo de una imagen medida. La reflectancia es dado por $R(x,y,\lambda)$, donde x e y son las coordenadas espaciales y λ es la longitud de onda.

$$R(x, y, \lambda) = \frac{I_{blanco}(x, y, \lambda) - I_m(x, y, \lambda)}{I_{blanco}(x, y, \lambda) - I_{oscuro}(x, y, \lambda)} \quad (15)$$

La imagen del patrón de referencia (teflón) tiene que tener la reflectancia homogénea en sus 168 bandas, esto asegura la reproducibilidad los experimentos. La fuente de luz tiene gran importancia en el momento de la calibración del sistema y de la adquisición de las imágenes, deben ser lo más estable posible en el transcurso del tiempo e iluminar el área a analizar de forma homogénea. En la figura 3.5 se observa el patrón de referencia iluminado por la fuente de luz.

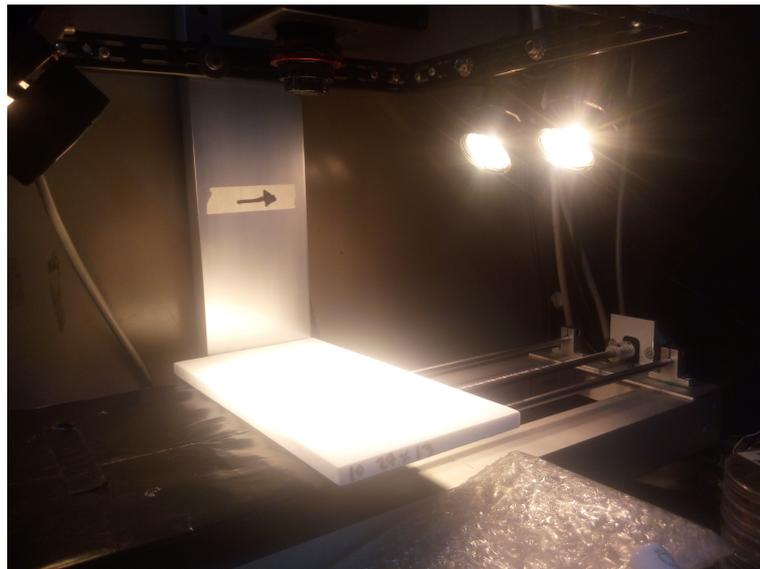


Figura 3.5: Patrón de referencia.

En la figura 3.6 se observan los espectros medios de todos los píxeles del teflón iluminado con la lámpara halógena, la cual fueron tomadas en dos diferentes días, con las luces del sistema y luces propias del ambiente de laboratorio, ambos espectros tienen un día de diferencia y se adquirieron minutos después de la calibración de rutina, el espectro azul y anaranjado presentan la característica de tener caídas de reflectancia en las primeras bandas y últimas; también presentan picos en las bandas 1020, 1140, 1380 y 1560 nm.

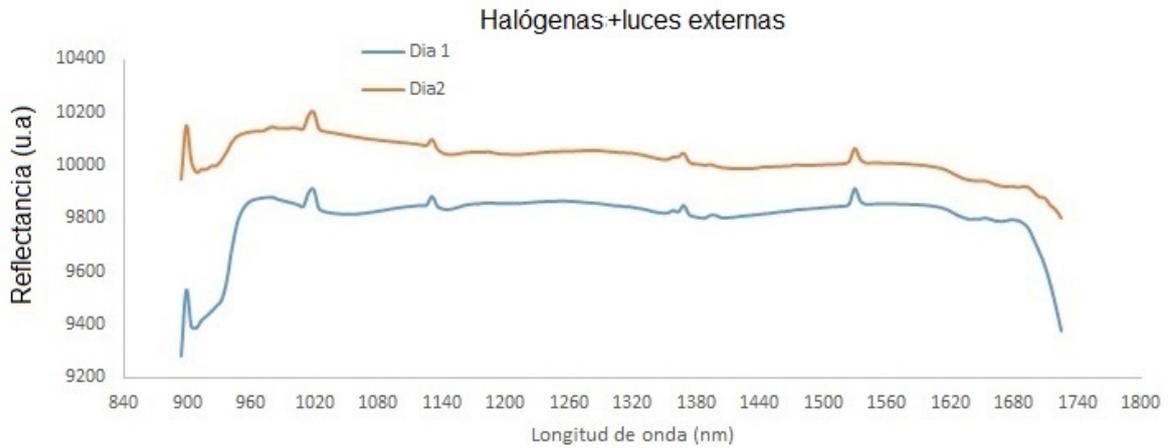


Figura 3.6: Espectro de teflón (reflectancia en unidades arbitrarias) con fuente luminosa de lámpara de tungsteno , monitor de PC y luz fluorescente.

También se registraron espectros medios de todos los píxeles del teflón en un ambiente que se aísla de las luces externas, cuya única fuente de luz es la del sistema HSI (lámpara de tungsteno). El tiempo en la cual se adquirieron fue segundos después de la calibración para ambos espectros y el tiempo que los separa es de un día (ver figura 3.7). En ambos espectros se mantiene la caída de reflectancia para las bandas iniciales y finales. También se observa que las reflectancias varían entre 10000 y 9900 (no está normalizado) la cual llega a ser aproximadamente $\sim 1\%$ de variación de reflectancia para sus respectivas bandas.

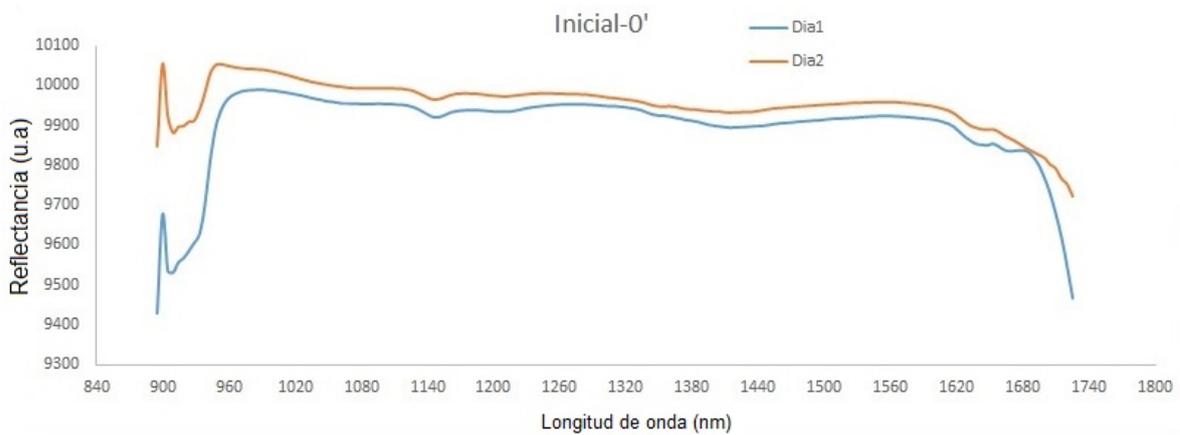


Figura 3.7: Espectro de teflón (reflectancia en unidades arbitrarias) en un ambiente aislado de luces externas y con fuente luminosa halógena.

Para continuar el análisis de la respuesta espectral del teflón, se analiza los espectros 5 minutos luego de la primera adquisición, en la figura 3.8 el espectro adquirido en el primer día cambia de forma entre 960 y 1140 nm, a pesar de dicho cambio la variación de reflectancia entre las bandas sigue siendo próximo a $\sim 1\%$ y cercano a 100000

unidades arbitraria de reflectancia.

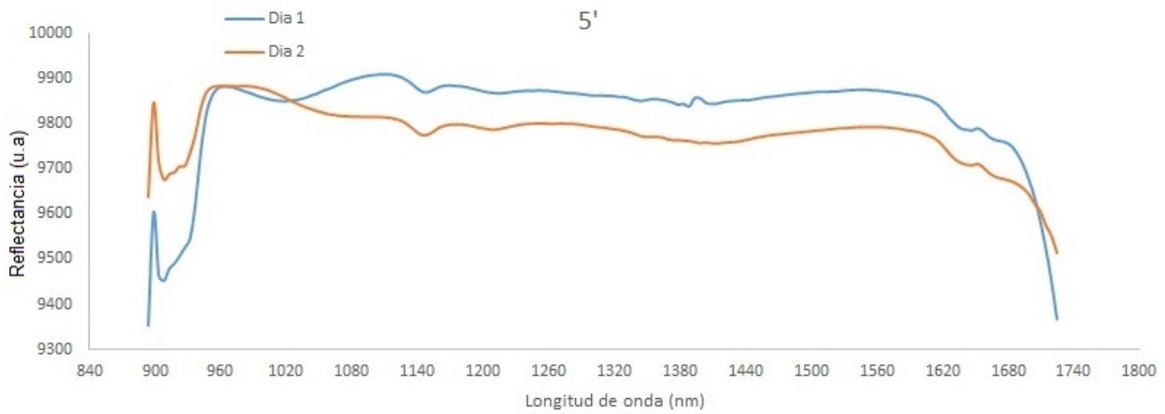


Figura 3.8: Espectro de teflón en un ambiente aislado y con fuente luminosa halógena con 5 minutos de diferencia entre un espectro y otro.

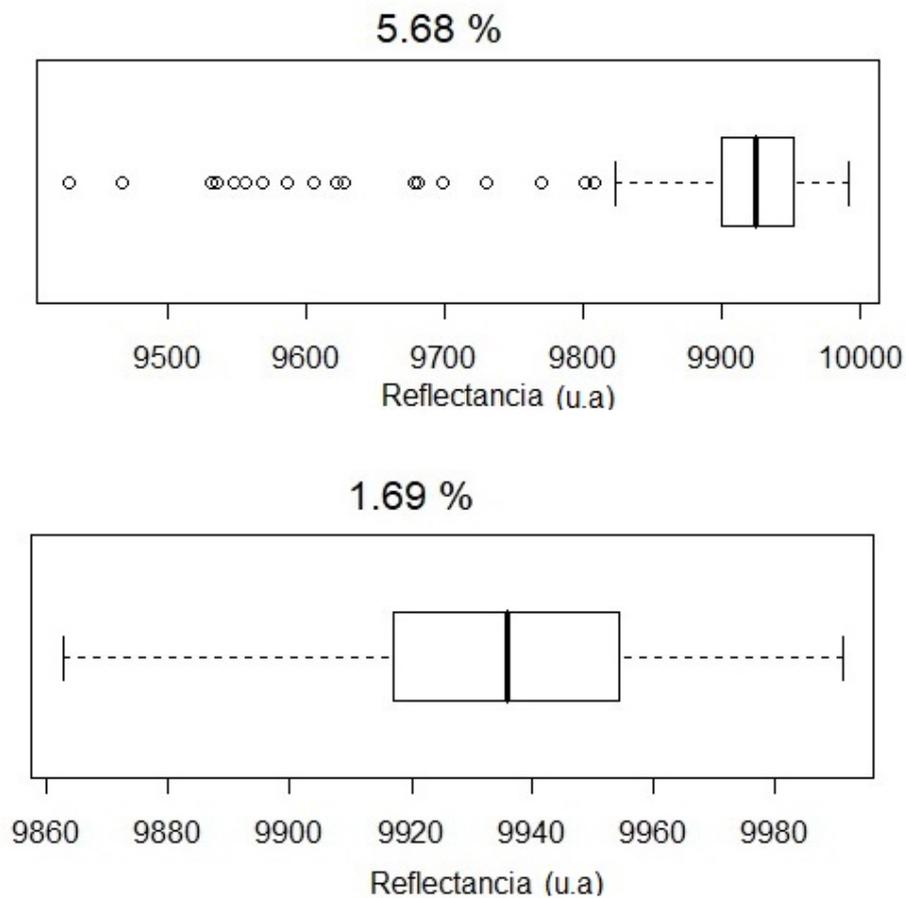


Figura 3.9: (PARTE SUPERIOR) Diagrama de cajas de datos de reflectancia de las 168 bandas con variación de 5.68 %, (PARTE INFERIOR) Diagrama de cajas de datos de reflectancia de las 136 bandas con variación de 1.69 %.

Las regiones de los espectros (bandas) que no dan resultados coherentes son elimi-

nadas y no son parte del preprocesamiento y procesamiento. Para ello calcularemos el rango de valores en la cual es adecuado considerar las bandas; Para ello se realiza un diagrama de caja como se observa en la parte superior de la figura 3.9. En el diagrama de caja se observa los valores anómalos que se ubican por debajo 9800 de reflectancia al considerar las 168 bandas y con una variación en porcentaje de 5.68 %, cuando no se toma los valores anómalos como se observa en el diagrama de caja de la parte inferior de la figura 3.9, el valor de la variación es de 1.69 % para las 140 bandas que están entre las bandas 12 a 151 correspondiente a las longitudes de 947 a 1647 nm respectivamente. La cantidad de bandas que se utilizan luego de este análisis es de 136 bandas.

3.2. Preparación de las muestras patrones de referencia (PR)

Las muestras a estudiar se clasificaron en dos grupos, el primer grupo formado por los patrones de referencia (PR) los cuales son muestras de suelos contaminados y un segundo grupo de muestras de suelo no contaminado.

Se utilizaron los siguientes instrumentos y materiales: Un horno o mufla para poder eliminar la mayor cantidad de humedad del suelo, un tamiz de 0.5 mm para poder tener las partículas del suelo de forma homogénea, una balanza para poder medir la masa del suelo que se mezclará con los DP y PR y diferentes materiales de laboratorio como: vasos precipitado, pipeta, embudo, varilla de vidrio y crisol.

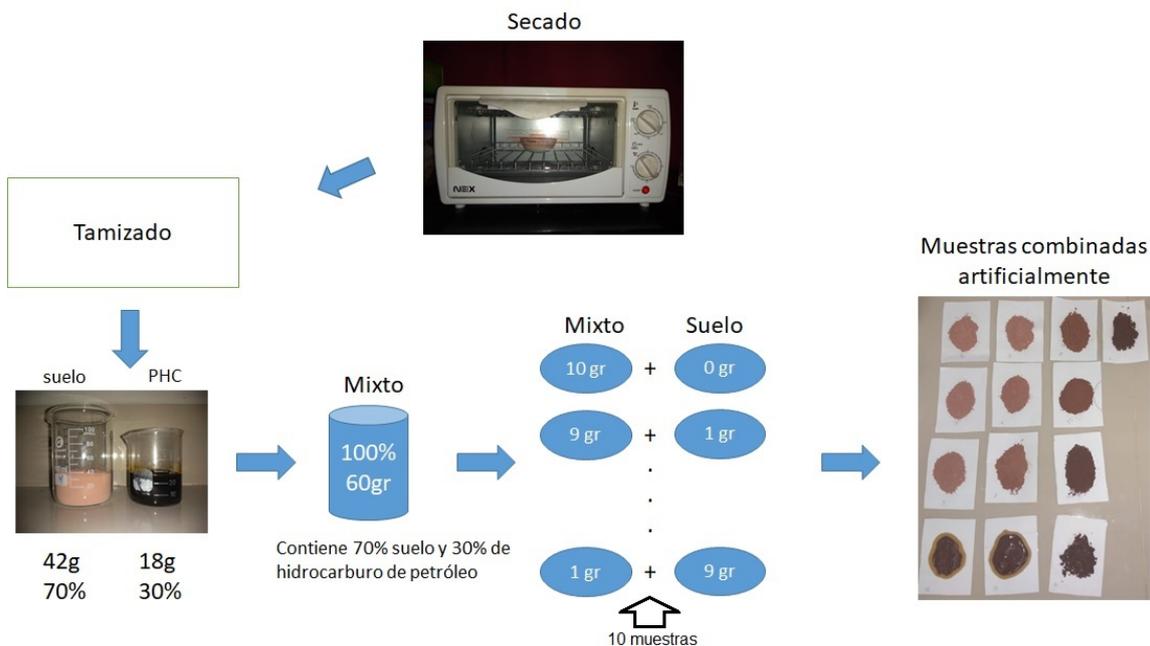


Figura 3.10: Esquema de la preparación de los patrones de referencia (PR) para los suelos mezclados con derivados de petróleo (DP).

3.2.1. Muestras de suelo y derivados de petróleo puros

Las muestras de suelo que se utilizaron se clasificaron según su textura en: suelo limoso, arcilloso y arenoso. Los contaminantes DP que se utilizaron son: crudo ligero y diésel obtenidos de diferentes ubicaciones (ver tabla 3.5).

Tabla 3.5: Muestras de suelos y derivados de petróleo puros indicando algunas características

Tipo	Muestras	Origen	Composición
suelos	Arcilloso	distribuidor local	
	Limoso	Cusco	
	Arenoso	distribuidor local	
Derivados de petróleo (DP)	Crudo ligero	Refinería de Piura	C, H, S y N
	Diésel	Distribuidor local	$C_{12}H_{26}$

En la parte inferior de la figura 3.11 se muestra los diferentes tipos de suelos: arenoso, arcilloso y limoso. En la parte superior el crudo de petróleo y diésel.



Figura 3.11: Suelos y derivados de petróleo puros utilizados para tener los patrones de referencia.

3.2.2. Patrones de referencia (PR)

Las muestras de los suelos se secaron en forma progresiva a $120^{\circ}C$ por un tiempo total de 1 hora en la cual se eliminó la mayor parte de humedad presente en los suelos, luego se tamizó con un tamiz de 0.5 mm. Con cada uno de los suelos se realizaron seis PR, los cuales son: de suelo limoso con crudo, limoso con diésel, arcilloso con crudo, arcilloso con diésel, arenoso con crudo y arenoso con diésel. Las concentraciones máximas (saturación del suelo) para cada tipo de suelo son: 70% de suelo y 30% de

contaminante (crudo de petróleo o diésel) para los suelos limosos y arcillosos y 80 % de suelo arenoso y 20 % de contaminante (crudo de petróleo o diésel). En la figura 3.12 se muestran dos patrones de suelos arcillosos con crudo y diésel respectivamente.



Figura 3.12: (IZQUIERDA) PR de 70 % de arcilla y 30 %pp-crudo de petróleo, (DERECHA) Patrón de referencia de 70 % de arcilla y 30 %pp-diésel.

De los seis PR de máxima concentración, los únicos que se reducirán la concentración en un intervalo de 3 % hasta llegar a la mínima concentración de 3 %pp – DP son los que tiene el suelo limoso y arcilloso, mientras los PR que tienen los suelos arenosos de concentración máxima de 20 %pp – DP se reducirán en un rango de 2 % hasta llegar a la mínima concentración de 2 %pp – DP, en total se tendrán 60 PR y tres suelos no contaminados (limoso, arcilloso y arenoso) que se utilizarán como blanco de referencia.

En la figura 3.13 se muestra 30 PR de diferentes concentraciones con crudo de petróleo y la figura 3.14 los PR con diésel. En cada figura se muestra: los patrones de los suelos arenosos de máxima y mínima concentración (parte superior), los patrones de los suelos limosos de máxima y mínima concentración (centro) y los patrones de los suelos arcillosos de máxima y mínima concentración (inferior).

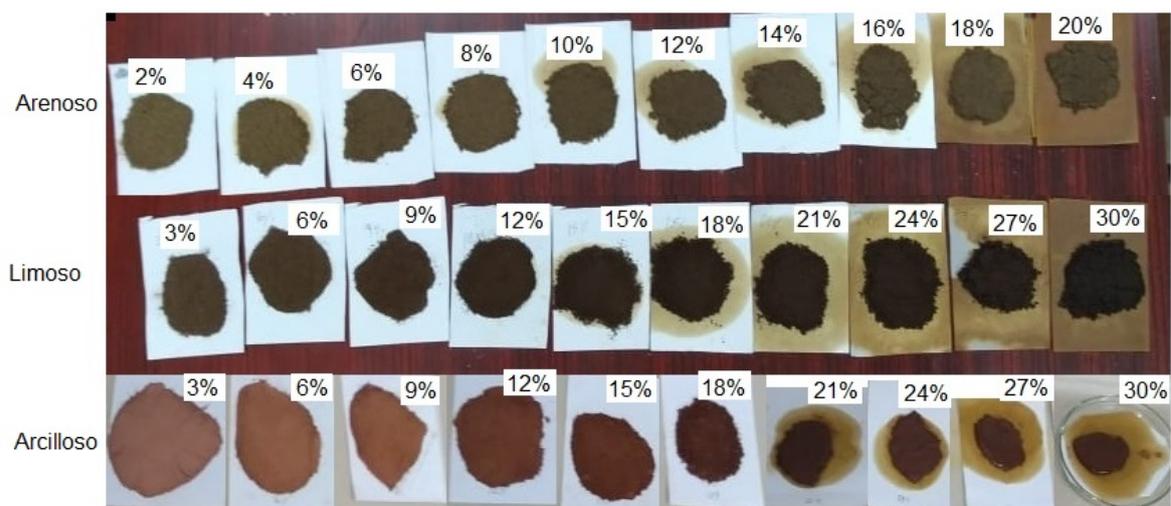


Figura 3.13: Patrones de referencia (PR) con crudo de petróleo, (SUPERIOR) suelos arenosos, (MEDIO) suelos limosos y (INFERIOR) suelos arcillosos.

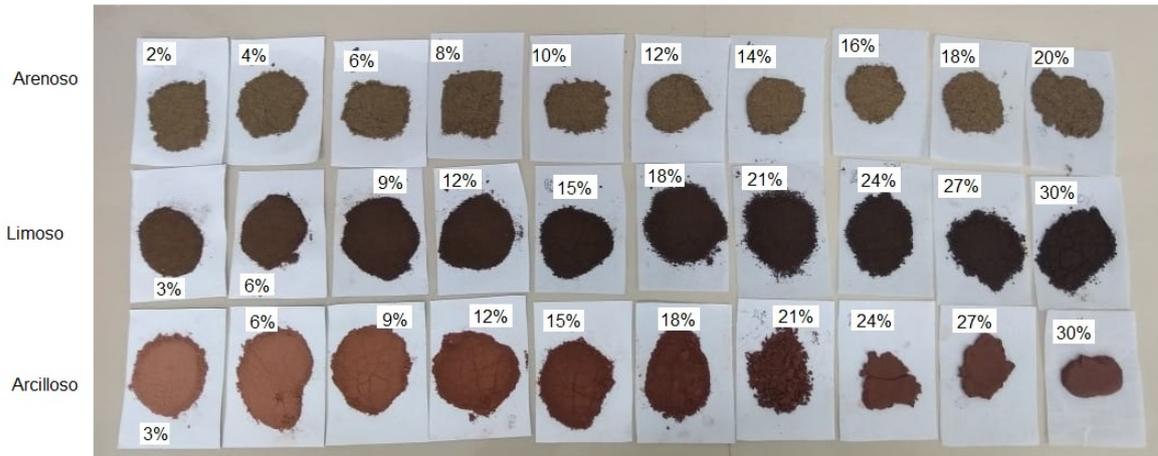


Figura 3.14: Patrones de referencia (PR) con diésel, (SUPERIOR) suelos arenosos, (MEDIO) suelos limosos y (INFERIOR) suelos arcillosos.

3.3. Adquisición de imágenes

Se utilizó el software SpectronPro para la adquisición de imágenes. Como ejemplo de una adquisición se observa en la figura 3.15 un mosaico de una HSI (HSI-MOSAICO) en el cual se unió cinco HSI (HSI-1:5), cada uno tiene en la parte superior el mismo suelo limoso sin contaminar. Los espectros de cada una de estas muestras se pueden observar en la figura 3.16, donde las muestras de suelo limoso sin contaminar (blanco-limoso) no cambias de forma y tienen la misma reflectancia. Para el resto de los espectros de los PR van cambiando de acuerdo a las concentraciones.

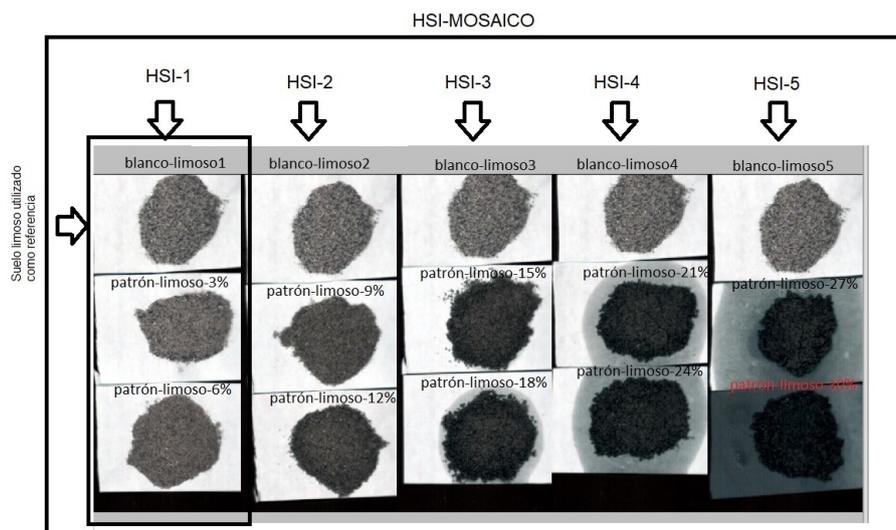


Figura 3.15: Mosaico de los patrones de referencia de los suelos limosos con crudo de petróleo y los suelos limosos sin contaminar. La fila superior corresponde al mismo suelo utilizado como referencia en cada adquisición.

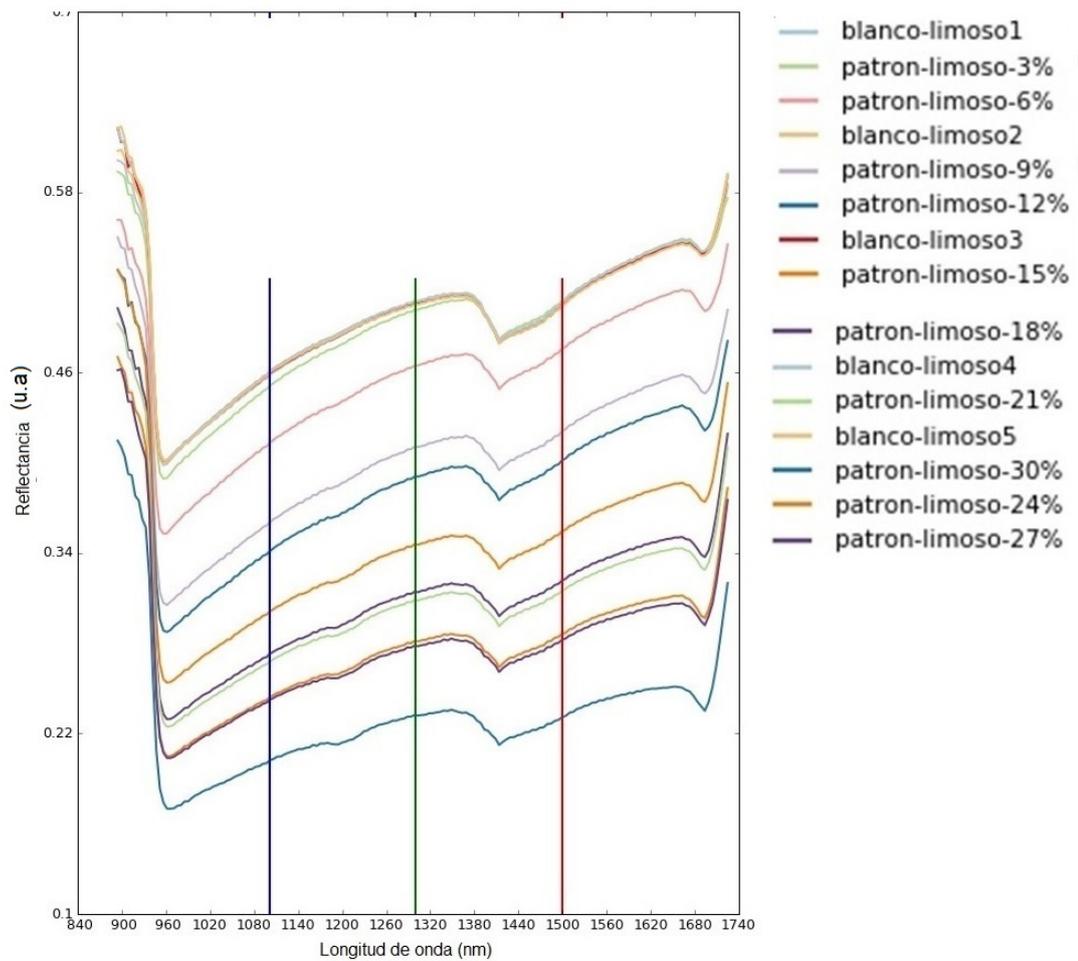


Figura 3.16: Espectros de los patrones de referencia de los suelo limosos con crudo de petróleo y los suelos limosos sin contaminar.

Las salidas de cada imagen es por medio de dos archivos: *.bil* y su meta data *hdr.bil*. Los cuales contienen información de la imagen hiperespectral. Se trata cada muestra (blancos y patrones de referencia) por separado en el software R-studio. Las dimensiones con las que se trabajan son de 200 píxeles x 200 píxeles y 136 bandas (ver el código ANEXO B).

3.4. Preprocesamiento

Para el preprocesamiento se realizan los métodos que se resumen en la figura 3.17.

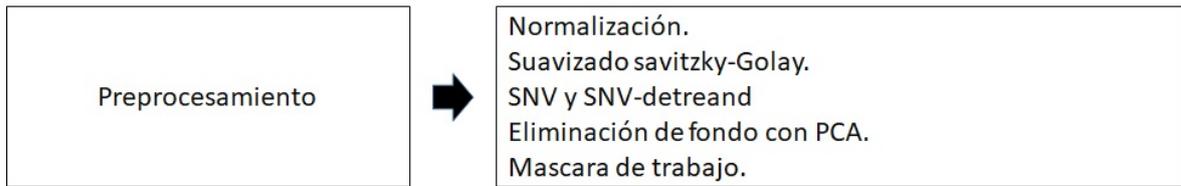


Figura 3.17: Partes del preprocesamiento.

Las HSI se trabajaron entre las longitudes de onda de 961 a 1632 nm y los métodos que se aplicaron son: Normalización máxima, Savitzky-Golay, Primera derivada, segunda derivada, SNV y de-trend y PCA. Como ejemplo de los métodos aplicados para el preprocesamiento se tomará la muestra del suelo arcilloso sin contaminar.

3.4.1. Normalización de la matriz del hipercubo

Aplicando la normalización en el lenguaje R. Se observa en la figura 3.18 una imagen de ejemplo de la banda 45 (de las 136 bandas) con sus valores normalizados, al lado de la imagen de la banda 45 se obtiene los espectros de los 100 primeros píxeles de la primera fila, los cuales son parte del fondo y de la muestra.

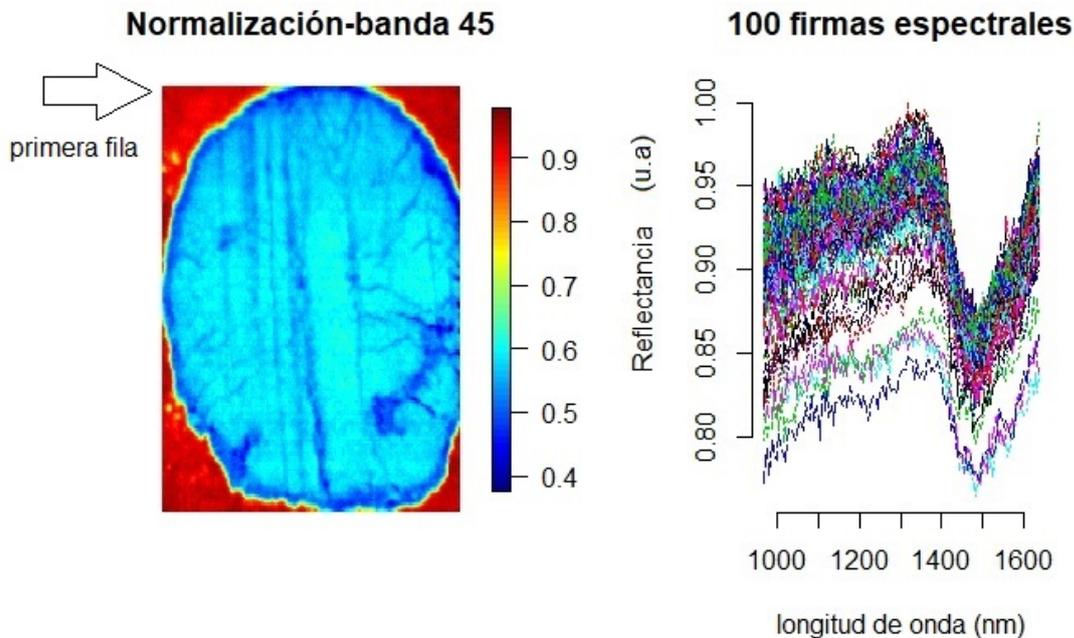


Figura 3.18: (IZQUIERDA) Imagen normalizada en la banda 45 del suelo arcilloso, (DERECHA) Espectros de las 100 primeros píxeles que conforman la imagen hiperespectral de 40 000 píxeles.

3.4.2. Suavizado Savitzky-Golay, primera y segunda derivada

Se sigue tomando como ejemplo el suelo arcilloso sin contaminar. Después de normalizar la HSI se realiza el suavizado Savitzky-Golay con polinomio de grado 5 y con tres diferentes cantidades de puntos 11, 25 y 47 de las bandas para la interpolación. Se busca que el suavizado contribuya en eliminar el ruido como también tener una mejor respuesta en la señal de la primera y segunda derivada (ver figura 3.19). Para ello obtenemos los resultados siguientes:

- Se observa que a medida que aumentamos la cantidad de puntos (11, 25 y 47) los espectros se suavizan más (figura 3.19).
- La región entre 1400 y 1600 se observa un pico de absorción el cual no cambia a medida que aumentamos la cantidad de puntos (figura 3.19).

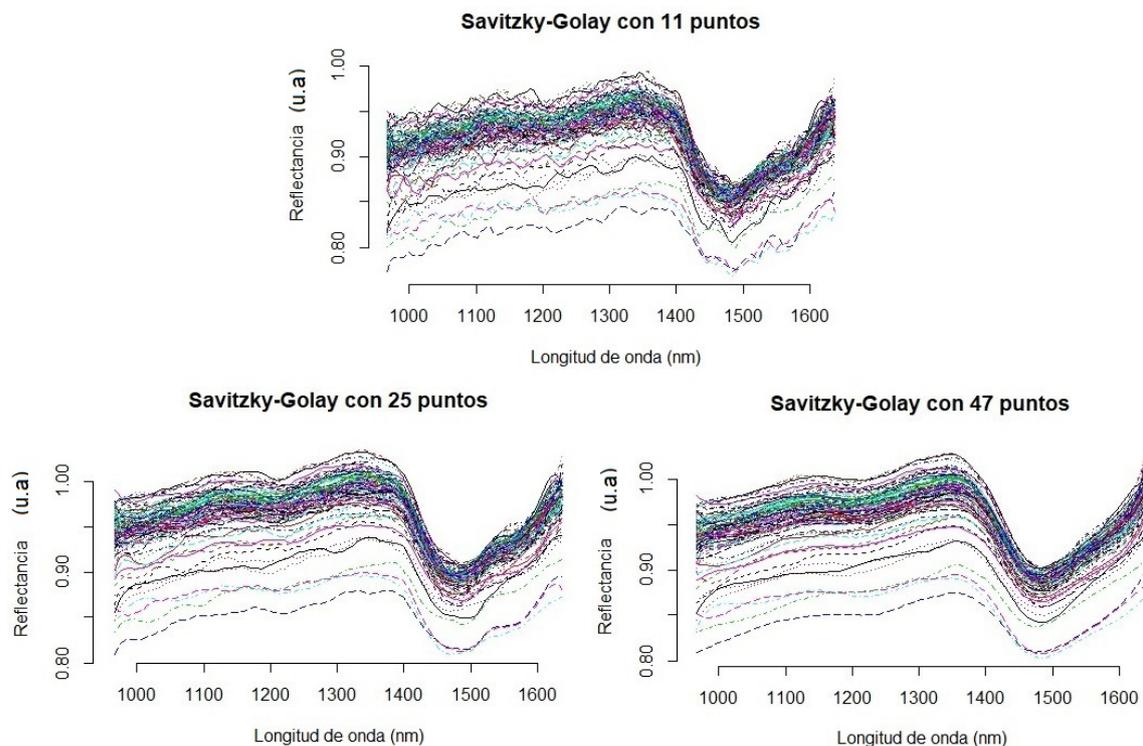


Figura 3.19: Savitzky-Golay: (SUPERIOR) polinomio de grado 5 con 11 puntos, (INFERIOR-IZQUIERDA) polinomio de grado 5 con 25 puntos, (INFERIOR-DERECHA) polinomio de grado 5 con 47 puntos.

- Con la primera derivada (figura 3.20) se observa que los espectros tienen la misma línea base, independiente de la cantidad de puntos que se analizan.
- Existe mayor señal en la primera derivada de 47 puntos que una con 11 puntos (figura 3.20).

- Se observa que a menor cantidad de punto es más difícil encontrar una región con una señal que se diferencie a las otras (figura 3.20).

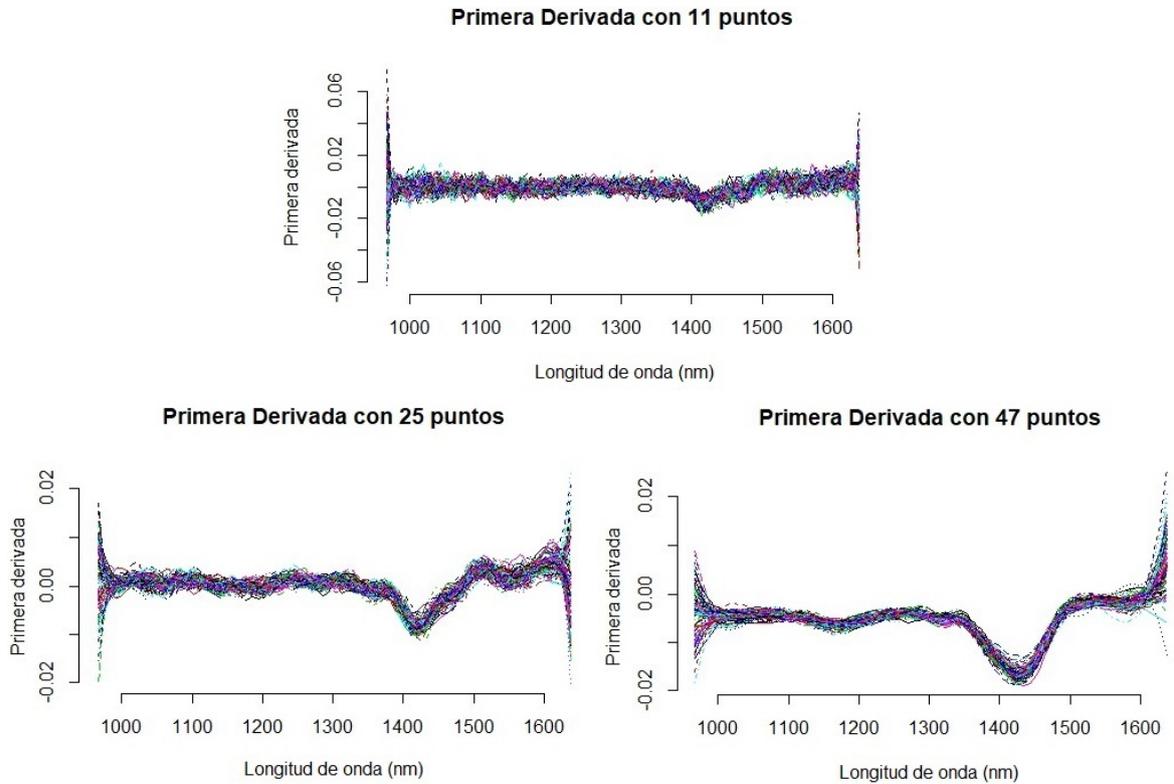


Figura 3.20: Primera derivada: (SUPERIOR) usando 11 puntos se observa una señal entre 1400 a 1500 nm, (INFERIOR-IZQUIERDA) usando 25 puntos se observa una mayor respuesta en la señal de 1400 a 1500 nm, (INFERIOR-DERECHA) con 47 puntos se observa una respuesta a la señal entre 1300 a 1400 nm.

- Con respecto a la segunda derivada (figura 3.21) se observa que para 47 puntos la señal en la región entre 1400 a 1600 nm es más intensa que los otros puntos.

Aplicando Savitzky-Golay para 47 puntos podemos ver que encontramos mayor respuesta en las bandas comparado con los otros puntos (11 y 25); por lo tanto, se aplicara savitzky-Golay de polinomio 5 y 47 puntos con la segunda derivada para relacionar la concentración y respuesta espectral.

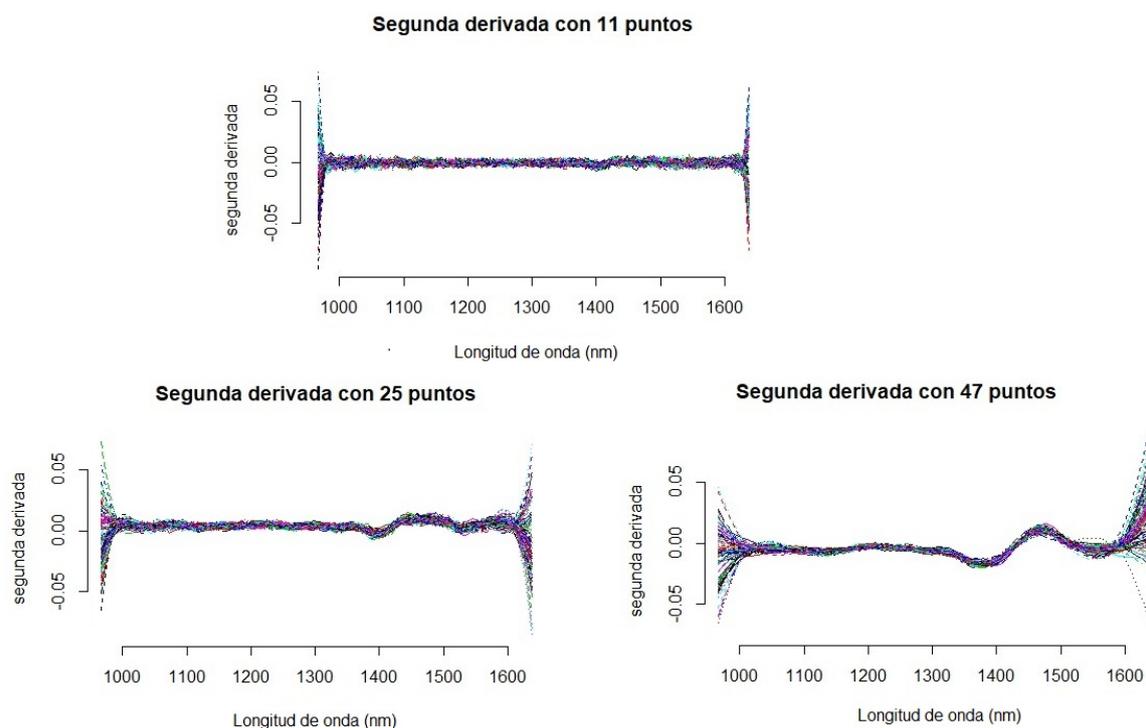


Figura 3.21: Segunda derivada: (SUPERIOR) usando 11 puntos se observa que se pierde la señal, (INFERIOR-IZQUIERDA) usando 25 puntos se observa una señal de 1300 a 1500 nm, (INFERIOR-DERECHA) con 47 puntos se observa un pico máximo en 1450 nm.

3.4.3. SNV y de-trend

Después de la normalización y el suavizado Savitzky-Golay sin derivada en el suelo arcilloso, se aplica el SNV (figura 3.22) o SNV-D (figura 3.23), se observa lo siguiente:

- Comparando con los espectros de la figura 3.19 y el de la figura 3.22 se observa que la dispersión en la reflectancia de los espectros disminuye (elimina escala multiplicativa).
- Comparando los espectros de la figura 3.19, figura 3.22 con la figura 3.23 se observa que la dispersión en la reflectancia de los espectros disminuye, eso es debido a que se eliminó la escala multiplicativa y la línea base.
- Con respecto a las imágenes que se obtienen de realizar SNV o SNV-D, podemos observar que al escoger una banda (de las 136) obtendremos la imagen de todos los píxeles los cuales pueden diferenciarse o también puede ser que el preprocesamiento SNV o SNV-D hagan que los píxeles en esa banda sean similares. Esto lo podemos observar en la figura 3.24 y la figura 3.25.

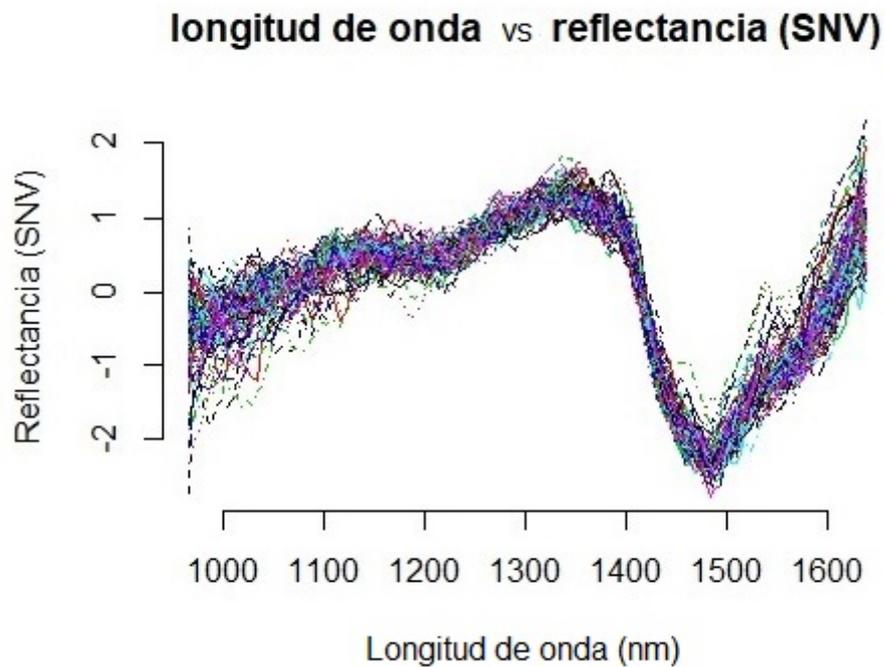


Figura 3.22: Espectros de los 100 primeros píxeles del suelo arcilloso a los cuales se aplicó Nor-max/Sv-G/SNV.

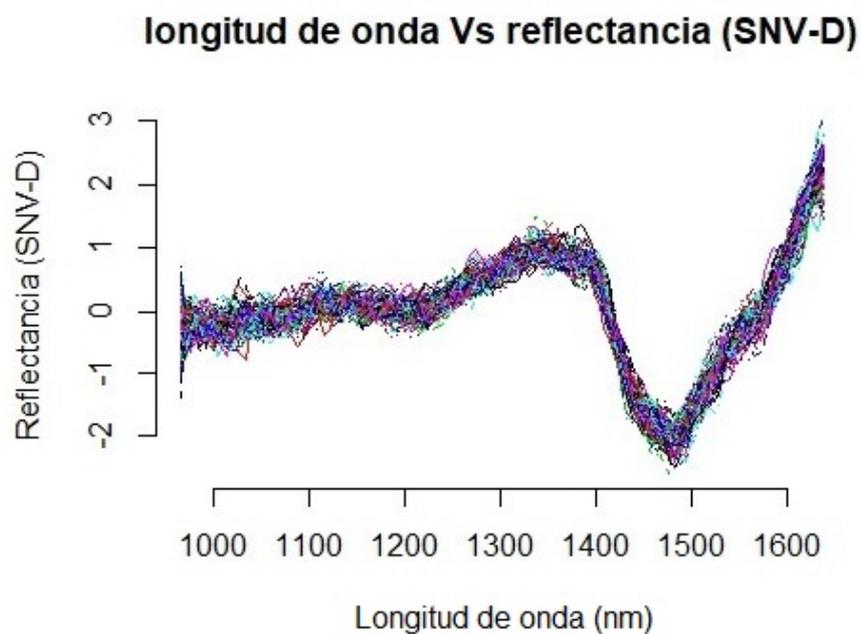


Figura 3.23: Espectros de los 100 primeros píxeles del suelo arcilloso a los cuales se aplicó Nor-max/Sv-G/SNV-D.

- En la figura 3.24 se muestra una imagen de la banda 30 del suelo arcilloso al

cual se aplicó Nor-max/Sv-G/SNV. La imagen muestra la diferencia en escala de grises entre la muestra y el fondo en comparación con la figura 3.24 y también se observa algunas características de la muestra, esto se debe que los valores de reflectancia SNV en los píxeles de la muestra y fondo son claramente diferentes.

SNV - banda 30 - 1102nm

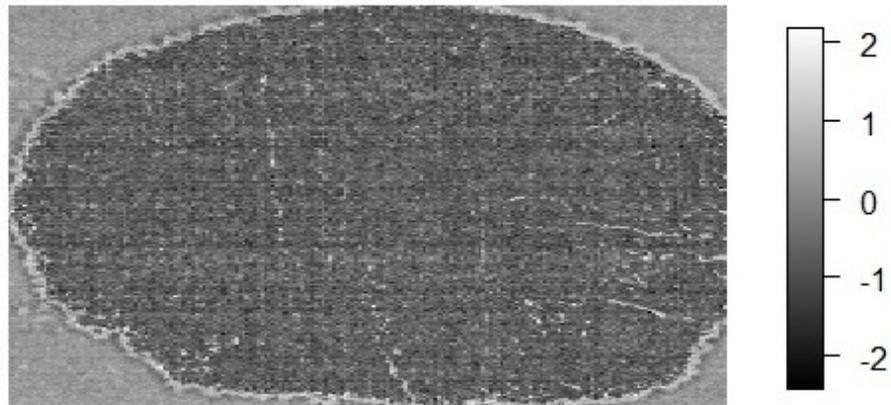


Figura 3.24: Imagen en escala de grises de la banda 30 del suelo arcilloso sin contaminar a la cual se aplicó Nor-max/Sv-G/SNV.

- En la figura 3.25 se muestra una imagen de la banda 30 del suelo arcilloso al cual se aplicó Nor-max/Sv-G/SNV-D. A diferencia de la imagen a la cual se aplicó SNV, los píxeles de esta imagen no se diferencian entre el suelo y la muestra. Esto se debe que al momento de eliminar la línea base algunas bandas de la muestra y del fondo coinciden.

SNV-D - banda 30 - 1102nm

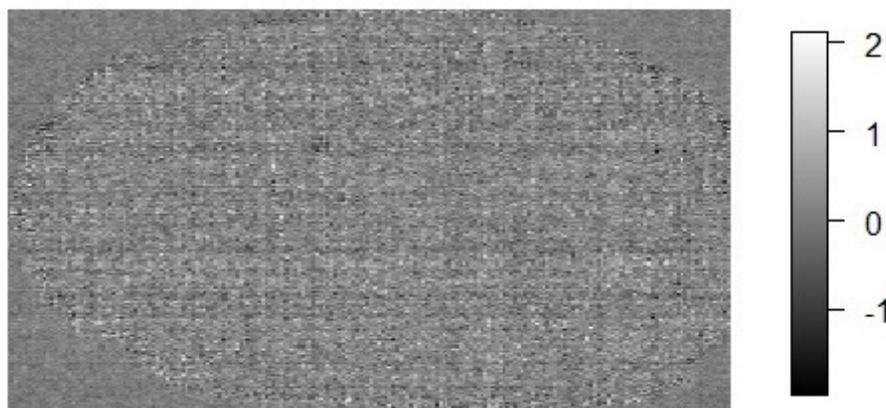


Figura 3.25: Imagen en escala de grises de la banda 30 del suelo arcilloso sin contaminar a la cual se aplicó Nor-max/Sv-G/SNV-D

Los preprocesamientos se van a dividir en diferentes combinaciones para diferentes fines, los cuales son: obtener una librería espectral propia (Nor-max/Sv-G(5,11,0)), diferenciar el fondo de la muestra (Nor-max/Sv-G(5,11,0)/SNV) y encontrar una relación entre la absorción de una banda y la concentración de los suelos contaminados (Nor-max/Sv-G(5,47,2)/SNV-D/2D). Se aplicarán las tres diferentes combinaciones (Nor-max/Sv-G (5,11,0)), (Nor-max/Sv-G(5,11,0)/SNV) y (Nor-max/Sv-G (5,47,2) /SNV-D/2D) para los suelos contaminados y los no contaminados.

3.4.4. Eliminación de fondo con PCA

Se realizará un análisis de componente principales (PCA) al preprocesamiento (Nor-max/Sv-G(5,11,0)/SNV) para todas las muestras con el fin de eliminar el fondo. Se toma como ejemplo la muestra de arcilla y se obtiene lo siguiente:

- En la figura 3.26 se observan los valores de la primera componente principal (Scores-máxima variabilidad) con variabilidad de 66.26% respecto a todos los píxeles en la nueva base, para el fondo los valores son mayores a 5 unidades (scores) y para la muestra son menores a 5 unidades (scores), esto se puede ver en el histograma de frecuencia. (derecha-figura 3.26).

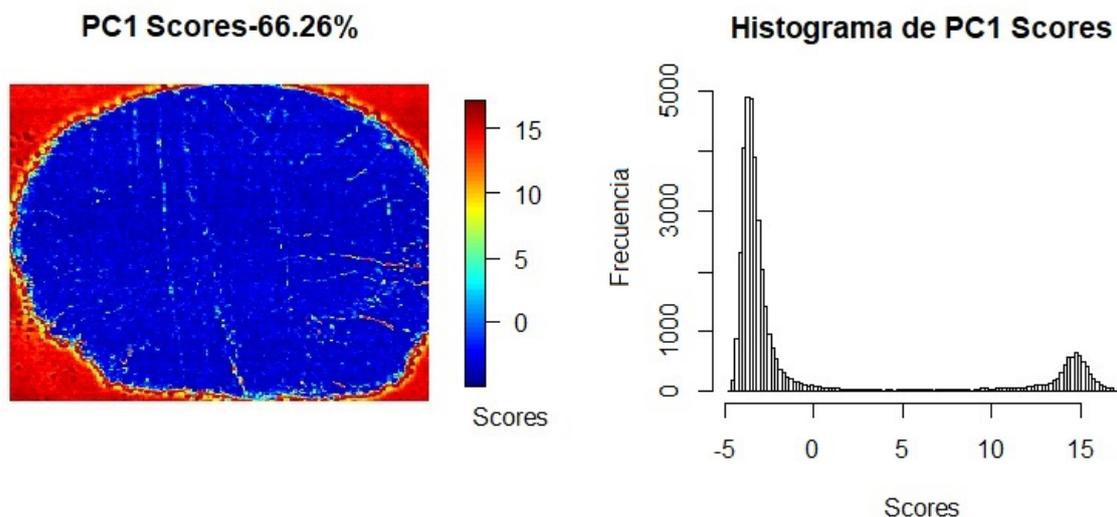


Figura 3.26: (IZQUIERDA) Scores de la primera componente de la muestra de suelo arcilloso, (DERECHA) histograma de los scores de la primera componente (N-max/Sv-G(5,11)/SNV/PC1).

- Se realiza la umbralización, haciendo que todos los scores mayores a 0 y menores a -5 sean igual a valores nulos (NA por las siglas en inglés *Not Available*), esto

nos permite tener las posiciones espaciales de los píxeles contenidos en la muestra (figura 3.26).

3.4.5. Identificación de la máscara de trabajo

Al ubicar las posiciones de cada píxel se puede realizar una máscara que cubre un área de nuestra imagen y que tiene como posición los píxeles de interés. A continuación se discute los píxeles que cubre dicha máscara.

- En la figura 3.27 superior se observa solo los píxeles que son de la muestra (de -5 a 0 scores), junto a ella está el histograma que le corresponde.
- Los valores de -5 a 0 scores son valores tentativos de la umbralización. Para conocer de manera más precisa los valores de umbralización se realiza una distribución de frecuencia por sturges (Solano and Álvarez, 2005) y se obtiene el rango que separa la muestra (valor inferior= $Q1-1.5*RI$ y valor superior= $Q3+1.5*RI$, donde RI es el rango intercuartílico, Q1 el primer cuartil y Q3 el tercer cuartil) y el fondo los cuales son -4.52 y -2.54. En la figura 3.27-inferior se observa los píxeles que se obtienen en ese rango y los valores anómalos superiores a -2.54 scores.

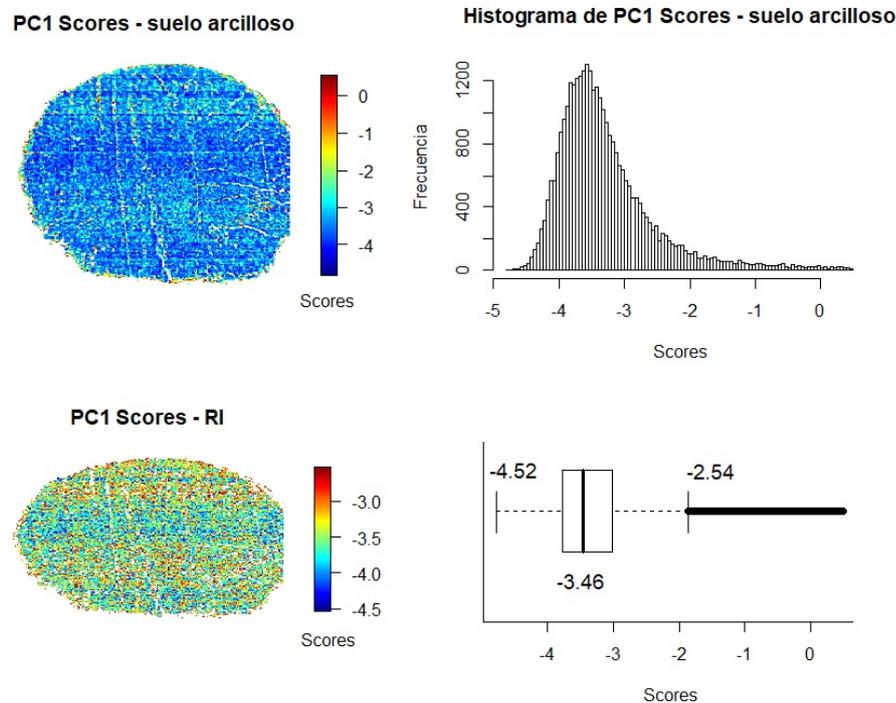


Figura 3.27: Parte superior de derecha a izquierda hacia abajo: Imagen umbralizada con, histograma de la imagen, imagen de los píxeles de valores entre -4.52 y -2.54 Scores del PC1 y diagrama de caja.

Se realiza este mismo proceso para conocer los rangos que contienen los píxeles de interés para el resto de muestras y de este modo realizar las máscaras para cada uno.

3.5. Procesamiento

Para el procesamiento se realizan los métodos que se resumen en la figura 3.28.

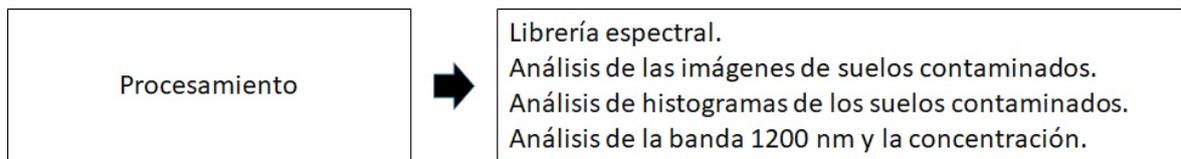


Figura 3.28: Partes del procesamiento.

Para el preprocesamiento se utiliza las máscaras encontradas anteriormente o los rangos de interés que se le aplicará a los siguientes preprocesamientos para los análisis correspondientes:

- Nor-max/Sv-G(5,11,0): Librería espectral, Análisis de las imágenes de suelos contaminados y análisis de histogramas de los suelos contaminados.
- Nor-max/Sv-G(5,47,2)/SNV-D/2D: Análisis de la banda de absorción a 1200 nm en función de la concentración del contaminante.

3.5.1. Construcción de librería espectral

Las muestras se dividieron en 6 grupos según el tipo de suelo (limoso, arcilloso y arenoso) y el contaminante que tiene (crudo y diésel). Cada HSI tiene un espectro de reflectancia medio que lo representa (reflectancia con unidades arbitrarias), logrando obtener 60 espectros correspondiente a los PR y 3 de los suelos sin contaminar los cuales pertenecerán a la librería espectral.

Los resultados que se obtienen de los espectros característicos de nuestra librería son:

- Limoso-crudo (figura 3.29): El espectro que presenta la mayor reflectancia para todas sus bandas es el suelo limoso sin contaminar, el resto de suelos limosos contaminado con crudo presentan espectros que se reducen a medida que el nivel de concentración aumenta, lo que se observa en todos los espectros es la banda de absorción típica del suelo en 1418 nm. Los espectros mostrados formarán parte de una librería espectral propia de suelo limoso contaminado con crudo de petróleo.
- Limoso-diésel (figura 3.30): El espectro limoso sin contaminar presenta mayor reflectancia en todas sus bandas, mientras que los suelos limosos contaminados

tiene una reflectancia que disminuye a medida que el nivel de contaminante aumenta. El suelo limoso contaminado con diésel a 3% tiene una menor reflectancia con respecto al suelo con crudo a 3%. De la misma manera que para el grupo de suelo limoso con crudo, este grupo formara parte de una librería espectral propia.

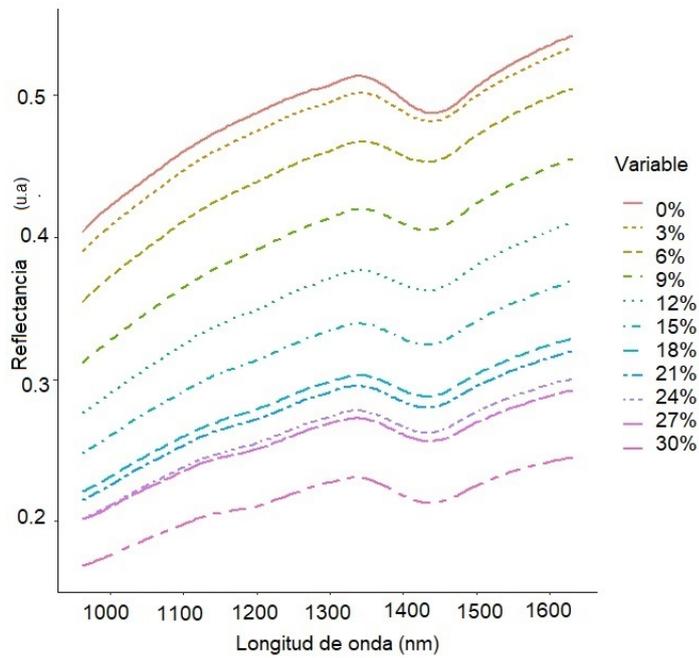


Figura 3.29: Espectros de reflectancia medios de las muestras de suelo limoso con diferentes concentraciones de crudo de petróleo.

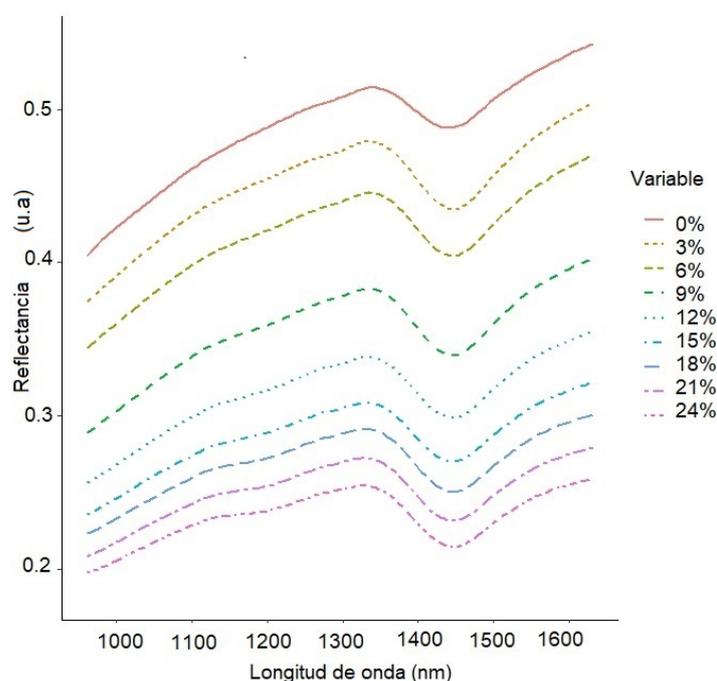


Figura 3.30: Espectros de reflectancia medios de las muestras de suelo limoso con diferentes concentraciones de diésel.

- Arcilloso-crudo (figura 3.31) y Arcilloso-diésel (figura 3.32): El espectro de suelo arcilloso sin contaminar es el que presenta mayor reflectancia. La reflectancia de los espectros de suelos contaminados disminuye a medida que la concentración aumenta. Los espectros que tienen una concentración mayor a 18 % comienzan a agruparse de manera desordenada.
- Arenoso-crudo (figura 3.33) y Arenoso-diésel (figura 3.34): La reflectancia máxima se da para el espectro de suelo arenoso mientras que para los suelos contaminados con crudo o diésel la reflectancia disminuye a medida que la concentración aumenta. Para el caso del espectro del suelo con 2 % de contaminante (mínima contaminación que se realizó) se observa que la reflectaría disminuyó de manera considerable.

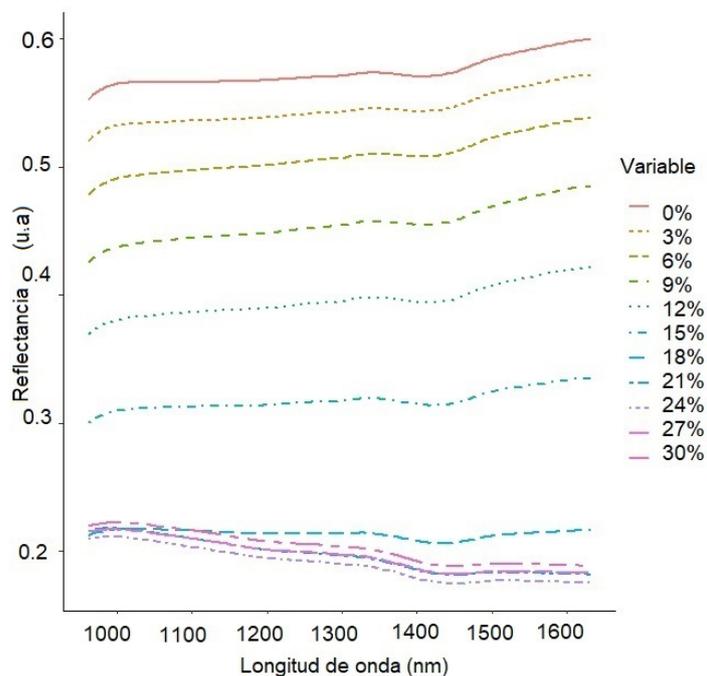


Figura 3.31: Espectros de reflectancia medios de las muestras de suelo arcilloso con diferentes concentraciones de crudo de petróleo.

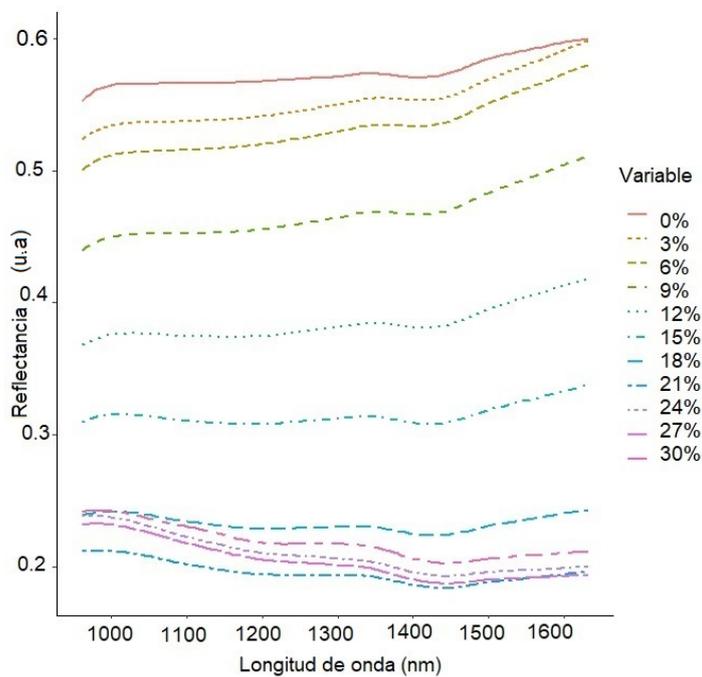


Figura 3.32: Espectros de reflectancia medios de las muestras de suelo arcilloso con diferentes concentraciones de diésel.

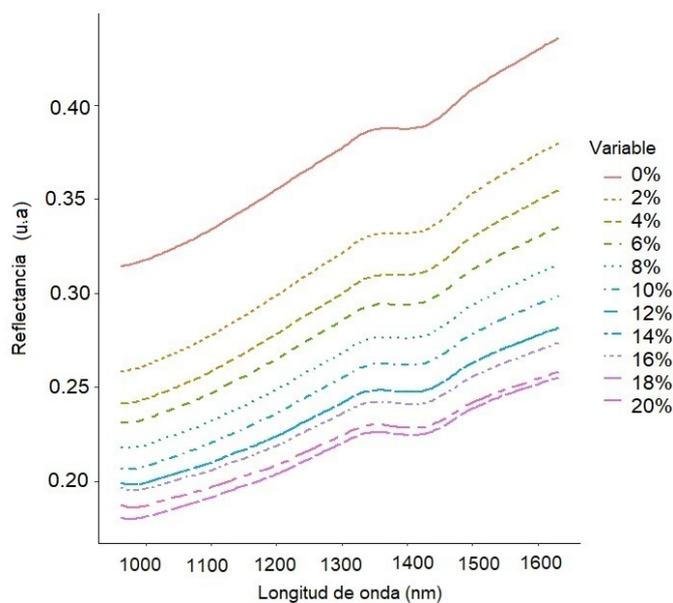


Figura 3.33: Espectros de reflectancia medios de las muestras de suelo arenoso con diferentes concentraciones de crudo de petróleo.

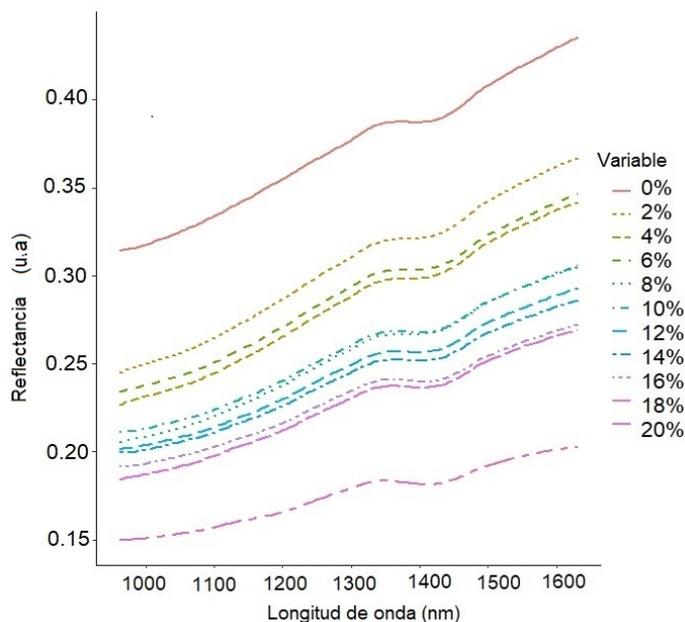


Figura 3.34: Espectros de reflectancia medios de las muestras de suelo arenoso con diferentes concentraciones de diésel.

Es de importancia poder conocer los espectros de nuestras muestras como también almacenarlas para un futuro uso. Es por ese motivo que todos los espectros propios formarán parte de nuestra librería espectral con las unidades de reflectancia en unidades arbitrarias (u.a), las cuales pudieran cambiarse a un valor real al multiplicar por el factor de 0.49 hallado en 3.1. Las unidades arbitrarias (u.a) de reflectancia no permite

ver de manera clara las bandas o regiones de absorción por tal motivo en 3.5.4 se utiliza el segundo procesamiento en los espectros.

3.5.2. Análisis de las imágenes de suelos contaminados

En esta parte recurrimos a la técnica de análisis de componentes principales y a la técnica de imágenes para formar mosaicos, se realizaron seis mosaicos (tabla 3.6) y cada mosaico contiene un suelo mezclado con un derivado de petróleo a diferentes niveles: mosaico 01 (muestras de suelo limoso con crudo de petróleo), mosaico 02 (muestras de suelo arcilloso con crudo de petróleo), mosaico 03 (muestras de suelo arenoso con crudo de petróleo), mosaico 04 (muestras de suelo limoso con diésel), mosaico 05 (muestras de suelo arcilloso con diésel) y mosaico 06 (muestras de suelo arenoso con diésel). A cada mosaico se realiza el análisis de componentes principales y se analiza solo el primer componente (PC1) por medio de los valores scores, todo los PC1 de los seis (6) mosaicos se juntaron en la figura 3.35 y sus resultados se discuten a continuación.

Tabla 3.6: Lista de mosaicos de los patrones de referencia

Mosaico	Suelo mezclado
01	Limoso con crudo
02	Arcilloso con crudo
03	Arenoso con crudo
04	Limoso con diésel
05	Arcilloso con diésel
06	Arenoso con diésel

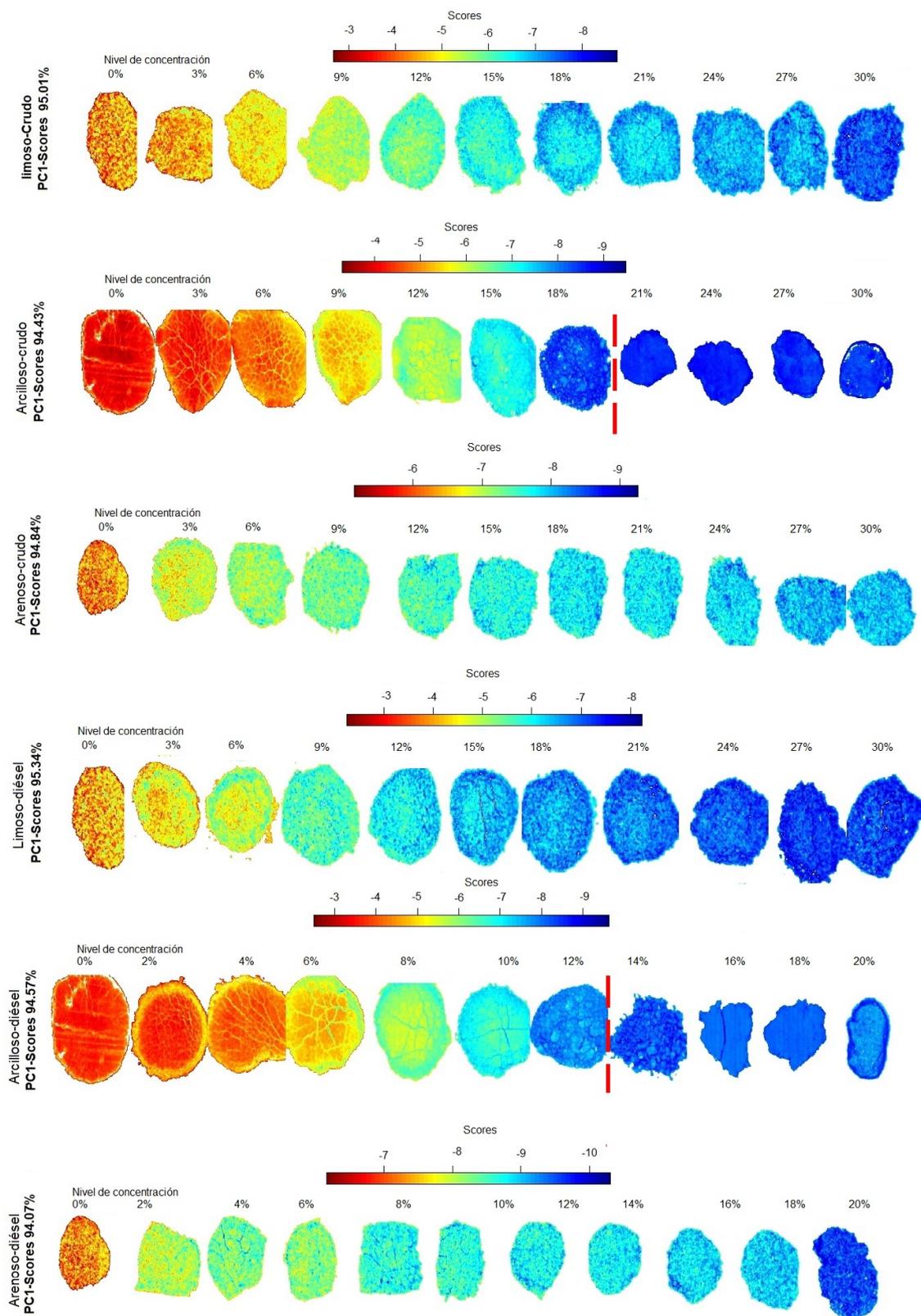


Figura 3.35: Imagen de las primeras componentes de los seis (6) mosaicos que muestran los dos tipos de suelos contaminado con diferentes derivados de petróleo a diferentes niveles.

Debido que el sistema de imagen hiperespectral es un sistema de clasificación rápida y en muchos casos instantánea, se buscó diferenciar los suelos con una escala de colores.

- Primer mosaico (limoso-crudo) y cuarto mosaico (limoso-diésel): En este caso el PC1 se obtuvo con una variabilidad de 95.01 % y 95.34 % para el crudo y diésel respectivamente. El suelo no contaminado tiene los valores más altos de scores y se le asigna el color rojo a medida que los scores disminuyen el color asignado es el azul.
- Segundo mosaico (arcilloso-crudo) y quinto mosaico (arcilloso-diésel): La variabilidad para el PC1 es 94.43 % y 94.57 % para el crudo y diésel respectivamente. El suelo arcilloso sin contaminar tiene el valor de scores más alto y por lo tanto se le asigna la escala de colores más alta (roja) a medida que la concentración aumenta a 3 % la escala de colores comienza a cambiar, logrando una diferencia para el suelo con 18 % de concentración y a valores más altos de concentración aparentemente la escala de colores no cambia.
- Tercer mosaico (arenoso-crudo) y sexto mosaico (arenoso-diésel): La variabilidad para el PC1 es 94.07 % y 95.34 % para el crudo y diésel respectivamente. Al suelo arenoso contiene diversos píxeles de diferentes valores en scores que se le asigna colores entre rojo y amarillo; pero a una concentración de 2 % los scores cambian de tal manera que se le asignan colores amarillos y verdes.

Para realizar un análisis más riguroso se realiza la prueba de hipótesis estadístico para poder conocer el error que se pudiera cometer al asignar una hipótesis nula (suelo sin contaminar) frente a una hipótesis alternativa (suelo contaminado).

3.5.3. Análisis del desplazamiento de los histogramas del componente principal (PC1)

Analizando los histogramas con sus respectivas funciones de distribución normal, se observa lo siguiente:

- La figura 3.36 muestra once (11) funciones de distribución normal del suelo limoso con diferentes niveles de concentración de crudo de petróleo, la curva de color negro presenta el nivel de concentración cero, esta curva nos sirve como blanco frente a las otras curvas a niveles superiores a cero, las curvas de mayor nivel de concentración se desplazan a la izquierda de la curva blanco.

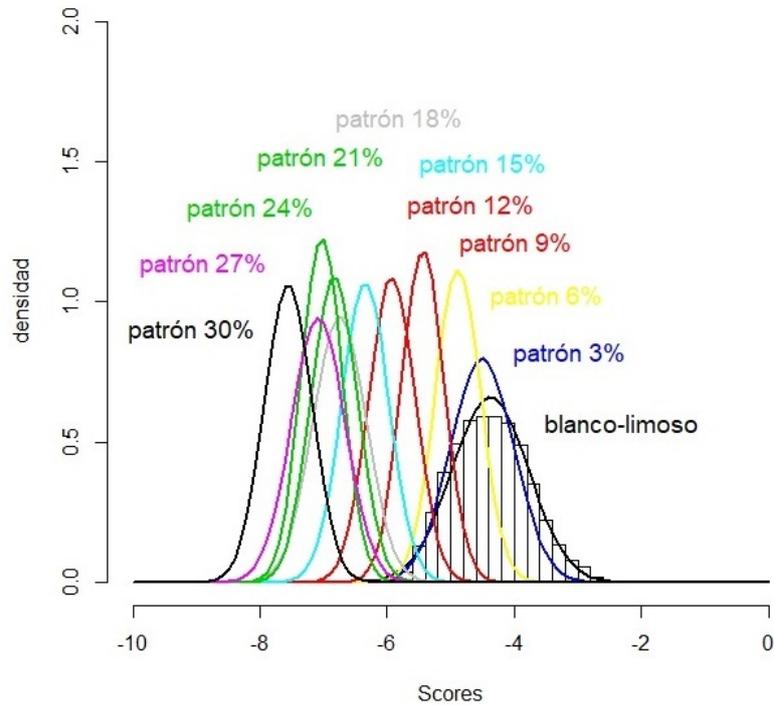


Figura 3.36: Funciones de distribución normal de los suelos limosos con crudo de petróleo a diferentes niveles de concentración frente al suelo limoso (blanco).

- La figura 3.37 muestra once (11) funciones de distribución normal del suelo arcilloso con diferentes niveles de concentración de crudo de petróleo, la curva de color negro presenta el nivel de concentración cero, en el suelo arcilloso las curvas con niveles de concentración mayores a 18% no se desplazan.
- La figura 3.38 muestra once (11) funciones de distribución normal del suelo arenoso con diferentes niveles de concentración de crudo de petróleo, la curva de color negro presenta el nivel de concentración cero, las curvas de diferentes niveles de concentración tienen un mayor desplazamiento en el suelo arenoso en comparación con el suelo limoso y arcilloso contaminado con crudo.

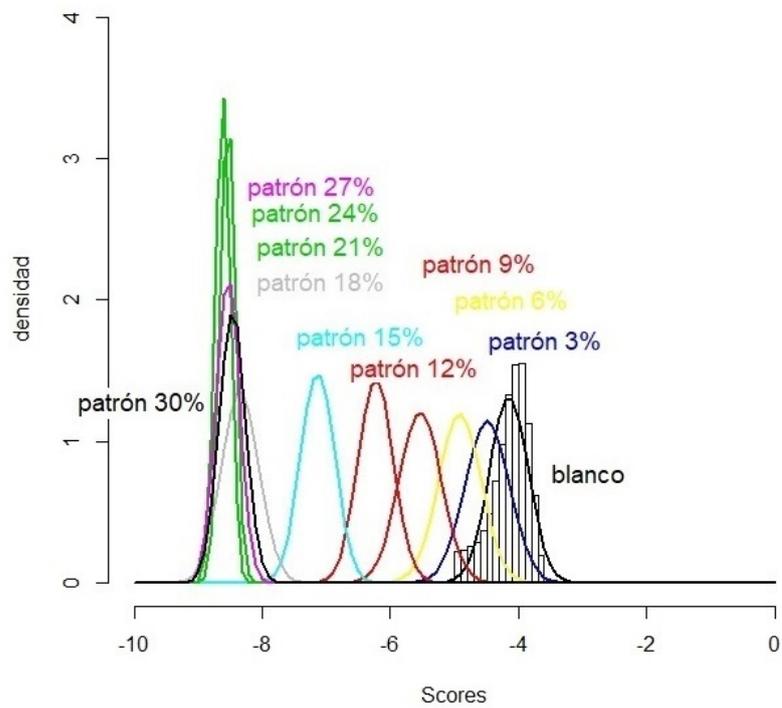


Figura 3.37: Funciones de distribución normal de los suelos arcillosos con crudo de petróleo a diferentes niveles de concentración frente al suelo arcilloso (blanco).

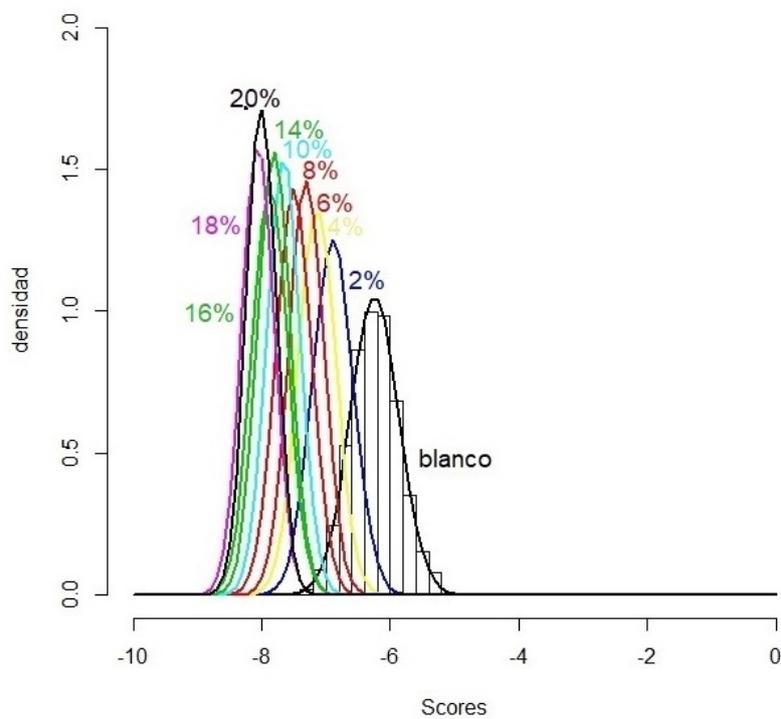


Figura 3.38: Funciones de distribución normal de los suelos arenosos con crudo de petróleo a diferentes niveles de concentración frente al suelo arenoso (blanco).

- La figura 3.39 muestra once (11) funciones de distribución normal del suelo li-

mosos con diferentes niveles de concentración de diésel, la curva negra del suelo limoso se tiene como referencia frente a los otros suelos que se desplazan a la izquierda a medida que aumenta la concentración.

- La figura 3.40 muestra once (11) funciones de distribución normal del suelo arcilloso con diferentes niveles de concentración de diésel, las curvas que tienen concentración mayor a 18 % dejan de desplazarse eso se refiere que el valor medio se mantiene y su desviación estándar disminuye.
- La figura 3.41 muestra once (11) funciones de distribución normal del suelo arenoso con diferentes niveles de concentración de diésel, al igual que el suelo arenoso con crudo, el desplazamiento es mayor con respecto a los otros suelos contaminados.
- Las tablas 3.7 y 3.8 contienen valores de interés obtenidos de la primera componente de cada suelo contaminado con derivados de petróleo y concentraciones diferentes en %pp-DP, las cuales son: media de la mezcla, desviación estándar de la mezcla (sd), distancia entre la media del suelo no contaminado y la media de la mezcla (D) donde el suelo sin contaminar se mantiene fijo, límite (limt) o nivel crítico entre la mezcla y la muestra no contaminada, probabilidad de clasificar un suelo no contaminado (prob suelo), probabilidad de clasificar la mezcla (prob mezcla), probabilidad de cometer el error tipo 1 (error tipo 1) y probabilidad de cometer el error tipo 2 (error tipo 2). Las probabilidades de clasificar el suelo y mezcla están relacionados con el error tipo 1 y tipo 2 respectivamente. Cuando la probabilidad de clasificar el suelo o mezcla es igual a 1 y el error tipo 1 o tipo 2 es igual a 0, todos los píxeles de una imagen se pueden clasificar.

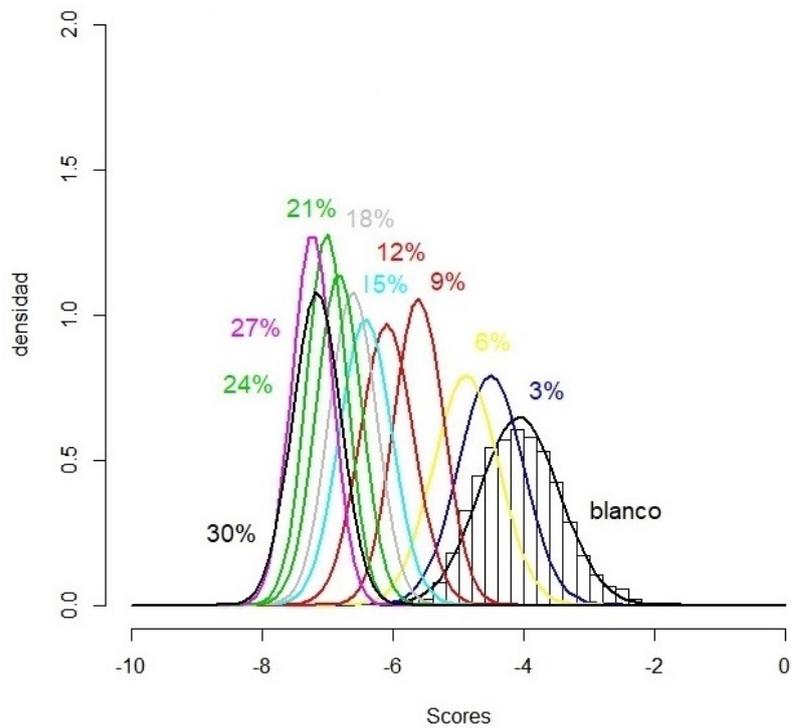


Figura 3.39: Función de densidad normal de los suelos limosos, arcillosos y arenosos contaminados con diésel frente al suelo no contaminado (blanco).

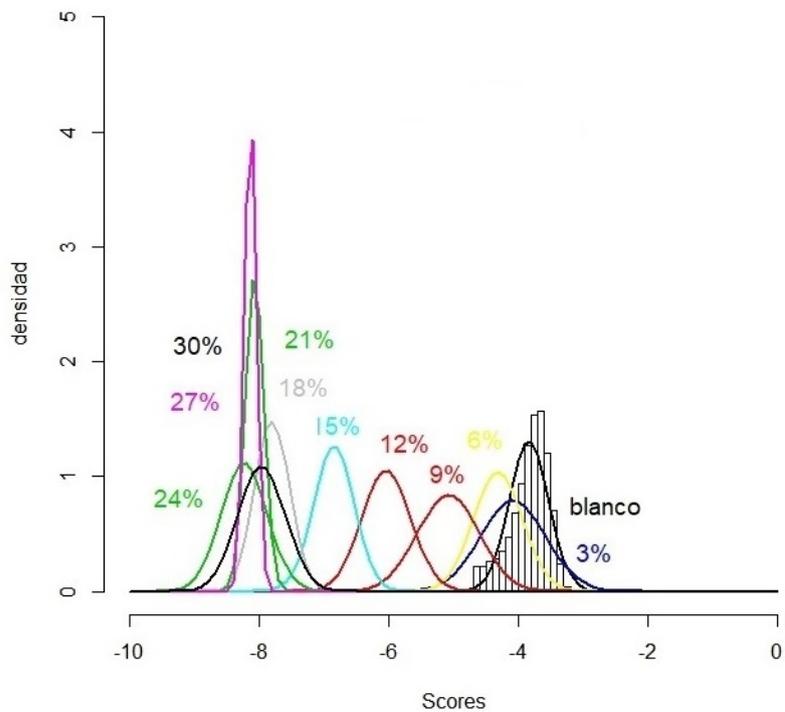


Figura 3.40: Función de densidad normal de los suelos limosos, arcillosos y arenosos contaminados con diésel frente al suelo no contaminado (blanco).

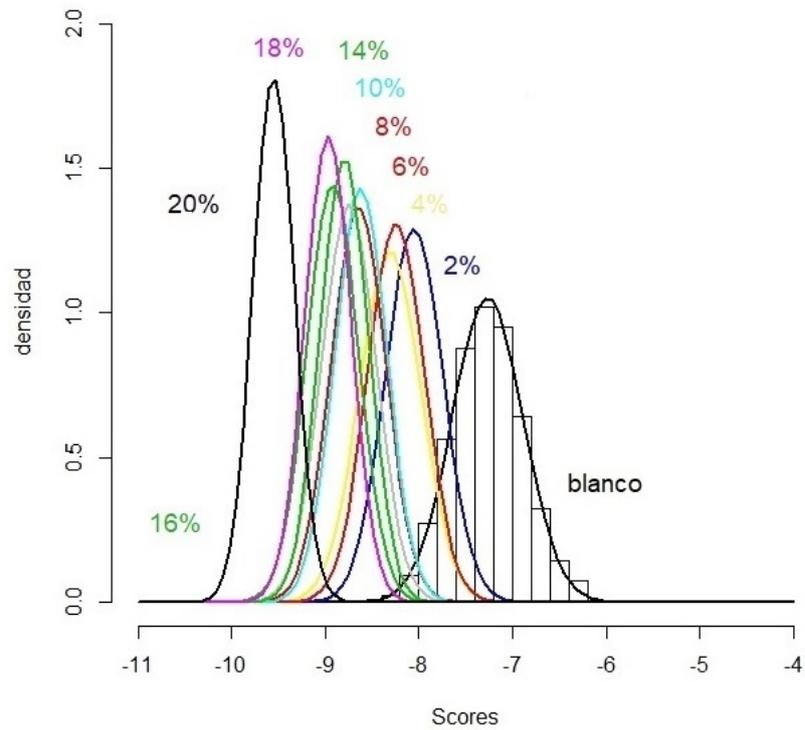


Figura 3.41: Función de densidad normal de los suelos limosos, arcillosos y arenosos contaminados con diésel frente al suelo no contaminado (blanco).

- Los valores de error tipo 1 y tipo 2 igual a cero (0) o cerca de él son recurrentes para concentraciones superiores como lo muestra las tablas 3.7 y 3.8.

Tabla 3.7: Valores obtenidos de los tres tipos de suelos contaminados con crudo de petróleo.

tipo	%pp DP	media mezcla	sd mezcla	D	limt	prob suelo	prob mezcla	error tipo 1	error tipo2
lim-cru	3	-4.50	0.50	0.13	-4.43	0.54	0.550	0.46	0.450
lim-cru	6	-4.88	0.36	0.51	-4.62	0.66	0.761	0.34	0.239
lim-cru	9	-5.43	0.34	1.06	-4.90	0.81	0.942	0.19	0.058
lim-cru	12	-5.92	0.37	1.55	-5.14	0.90	0.982	0.10	0.018
lim-cru	15	-6.34	0.37	1.97	-5.35	0.95	0.996	0.05	0.004
lim-cru	18	-6.74	0.42	2.37	-5.56	0.97	0.998	0.03	0.002
lim-cru	21	-6.83	0.37	2.46	-5.60	0.98	1.000	0.02	0.000
lim-cru	24	-7.02	0.33	2.65	-5.70	0.99	1.000	0.01	0.000
lim-cru	27	-7.08	0.42	2.71	-5.72	0.99	0.999	0.01	0.001
lim-cru	30	-7.55	0.37	3.18	-5.96	1.00	1.000	0.00	0.000
arc-cru	3	-4.431	0.347	0.336	-4.263	0.710	0.686	0.29	0.314
arc-cru	6	-4.860	0.335	0.764	-4.478	0.896	0.873	0.10	0.127
arc-cru	9	-5.477	0.330	1.382	-4.786	0.989	0.982	0.01	0.018
arc-cru	12	-6.171	0.277	2.075	-5.133	1.000	1.000	0.00	0.000
arc-cru	15	-7.078	0.268	2.983	-5.587	1.000	1.000	0.00	0.000
arc-cru	18	-8.289	0.293	4.193	-6.192	1.000	1.000	0.00	0.000
arc-cru	21	-8.484	0.119	4.389	-6.290	1.000	1.000	0.00	0.000
arc-cru	24	-8.560	0.115	4.465	-6.328	1.000	1.000	0.00	0.000
arc-cru	27	-8.480	0.185	4.384	-6.288	1.000	1.000	0.00	0.000
arc-cru	30	-8.410	0.208	4.314	-6.253	1.000	1.000	0.00	0.000
are-cru	2	-6.939	0.319	0.650	-6.614	0.804	0.846	0.20	0.154
are-cru	4	-7.185	0.294	0.896	-6.737	0.881	0.937	0.12	0.063
are-cru	6	-7.353	0.274	1.064	-6.821	0.919	0.974	0.08	0.026
are-cru	8	-7.548	0.279	1.259	-6.918	0.951	0.988	0.05	0.012
are-cru	10	-7.705	0.260	1.416	-6.997	0.969	0.997	0.03	0.003
are-cru	12	-7.854	0.256	1.565	-7.071	0.980	0.999	0.02	0.001
are-cru	14	-7.854	0.256	1.565	-7.071	0.980	0.999	0.02	0.001
are-cru	16	-7.919	0.278	1.630	-7.104	0.984	0.998	0.02	0.002
are-cru	18	-8.104	0.251	1.815	-7.196	0.992	1.000	0.01	0.000
are-cru	20	-8.052	0.232	1.764	-7.171	0.990	1.000	0.01	0.000

Tabla 3.8: Valores obtenidos de los tres tipos de suelos contaminados con diésel.

tipo	%pp	media	sd	D	limt	prob	prob	error	error
	DP	mezcla	mezcla			suelo	mezcla	tipo 1	tipo2
lim-die	3	-4.472	0.503	0.437	-4.253	0.639	0.668	0.36	0.332
lim-die	6	-4.850	0.503	0.816	-4.442	0.746	0.791	0.25	0.209
lim-die	9	-5.574	0.377	1.540	-4.804	0.895	0.979	0.11	0.021
lim-die	12	-6.059	0.412	2.025	-5.047	0.950	0.993	0.05	0.007
lim-die	15	-6.384	0.403	2.350	-5.209	0.972	0.998	0.03	0.002
lim-die	18	-6.583	0.369	2.549	-5.309	0.981	1.000	0.02	0.000
lim-die	21	-6.796	0.348	2.762	-5.415	0.988	1.000	0.01	0.000
lim-die	24	-6.986	0.311	2.952	-5.510	0.992	1.000	0.01	0.000
lim-die	27	-7.207	0.310	3.172	-5.620	0.995	1.000	0.00	0.000
lim-die	30	-7.135	0.369	3.101	-5.585	0.994	1.000	0.01	0.000
arc-die	3	-4.050	0.505	0.247	-3.926	0.658	0.597	0.34	0.403
arc-die	6	-4.286	0.385	0.483	-4.044	0.787	0.735	0.21	0.265
arc-die	9	-5.046	0.475	1.244	-4.424	0.980	0.905	0.02	0.095
arc-die	12	-6.008	0.379	2.206	-4.905	1.000	0.998	0.00	0.002
arc-die	15	-6.818	0.316	3.016	-5.310	1.000	1.000	0.00	0.000
arc-die	18	-7.772	0.269	3.970	-5.787	1.000	1.000	0.00	0.000
arc-die	21	-8.202	0.358	4.400	-6.002	1.000	1.000	0.00	0.000
arc-die	24	-8.038	0.145	4.236	-5.920	1.000	1.000	0.00	0.000
arc-die	27	-8.100	0.094	4.298	-5.951	1.000	1.000	0.00	0.000
arc-die	30	-7.934	0.366	4.132	-5.868	1.000	1.000	0.00	0.000
are-die	2	-8.103	0.309	0.784	-7.711	0.849	0.898	0.15	0.102
are-die	4	-8.357	0.329	1.038	-7.838	0.914	0.943	0.09	0.057
are-die	6	-8.295	0.303	0.976	-7.807	0.901	0.946	0.10	0.054
are-die	8	-8.698	0.291	1.379	-8.008	0.965	0.991	0.03	0.009
are-die	10	-8.671	0.278	1.353	-7.995	0.963	0.993	0.04	0.007
are-die	12	-8.798	0.290	1.480	-8.059	0.974	0.995	0.03	0.005
are-die	14	-8.849	0.259	1.530	-8.084	0.978	0.998	0.02	0.002
are-die	16	-8.972	0.276	1.653	-8.145	0.985	0.999	0.01	0.001
are-die	18	-9.020	0.247	1.702	-8.170	0.988	1.000	0.01	0.000
are-die	20	-9.602	0.219	2.283	-8.460	0.999	1.000	0.00	0.000

Con los valores de las tablas se realizaron las siguientes gráficas: En la figura 3.42 se muestra el cambio de la distancia o desplazamiento de los diferentes niveles de concentración del suelo contaminado frente al no contaminado, en la figura 3.43 los valores medios con su desviación estándar de los seis grupos que se analiza y en la figura 3.44 se muestra el cambio del error respecto a la concentración en $\%pp - DP$. De esos gráficos se obtiene los siguientes resultados:

- La desviación estándar de un suelo no contaminado es siempre mayor que un suelo contaminado (figura 3.42 y figura 3.43).
- Existe mayor dispersión en el suelo no contaminado respecto a su media y menor

dispersión en el suelo contaminado respecto a su media.

- La distancia entre los valores medios de suelo contaminado y no contaminado siempre aumenta a medida que la concentración aumenta.
- Para los suelos arcillosos, la distancia entre el suelo contaminado y no contaminado llega a un límite en las concentraciones mayores a 18 %pp-DP (figura 3.42).

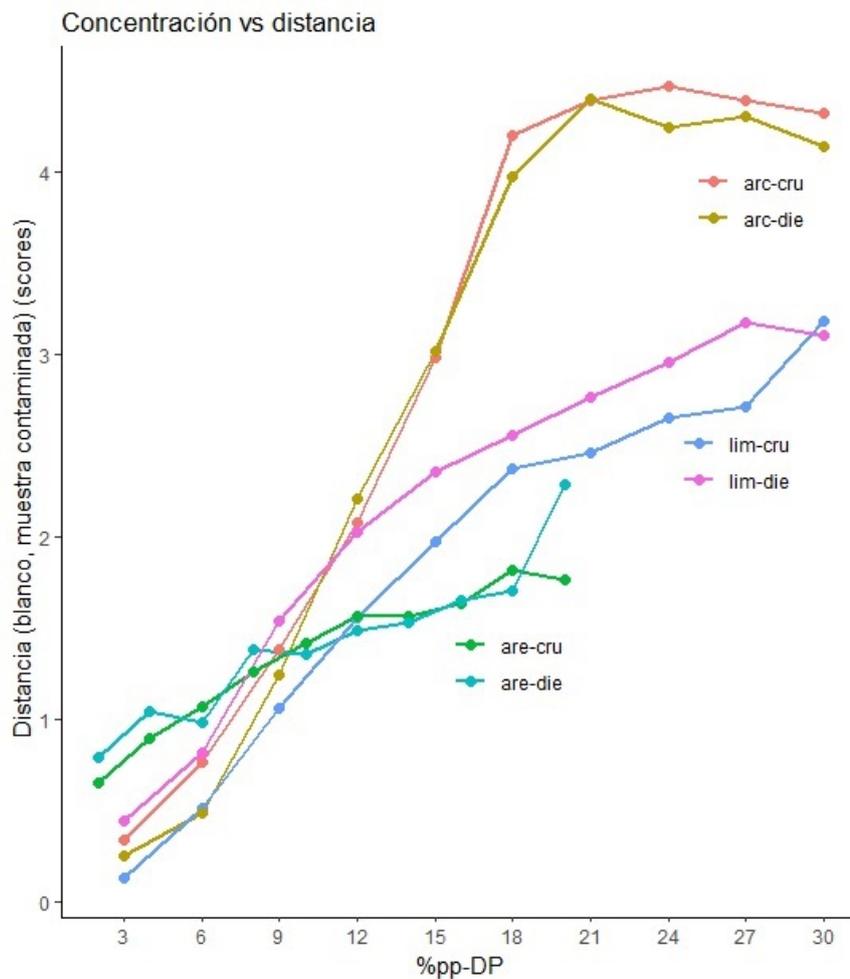


Figura 3.42: Distancia de los valores medios del blanco o muestra no contaminada a las muestras contaminadas o patrones de referencia.

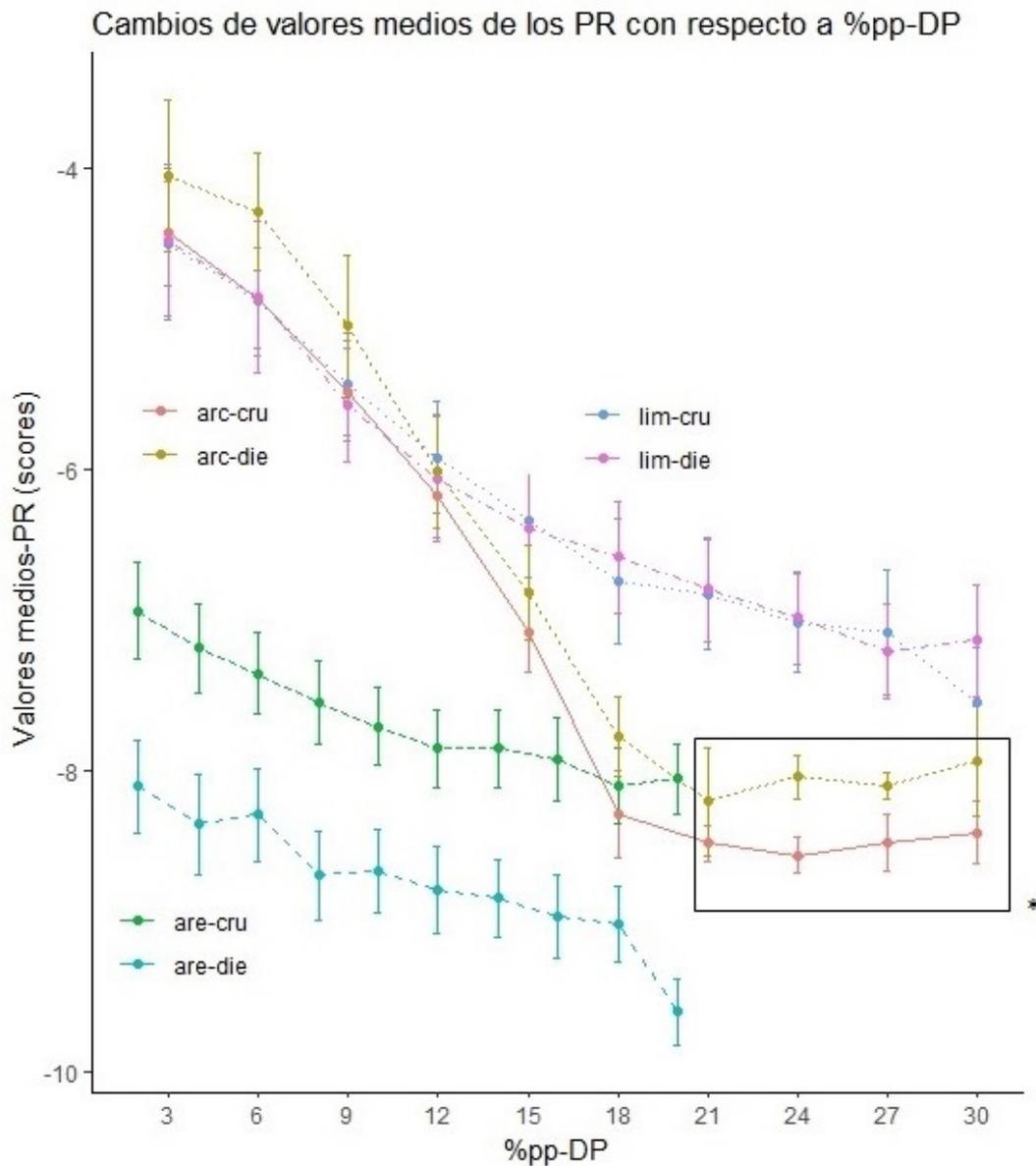


Figura 3.43: Valores medios de las muestras contaminadas o patrones de referencia con su desviación estándar (sd).

Con respecto a la probabilidad de cometer errores de tipo 1 y tipo 2 en la figura 3.44 se observa que los suelos arenosos contaminados son los que tienen menor probabilidad de cometer el error de tipo 1 y tipo 2 a concentraciones bajas. Estas probabilidades de cometer estos errores disminuyen a medida que la concentración aumenta para los 6 diferentes tipos de mezclas, se observa que estas probabilidades son:

- suelo limoso-crudo: (probabilidad del error tipo 1 o tipo 2) < 5% se da para concentraciones mayores a 9%pp-DP y estas probabilidades son nulas para valores de concentración igual o mayores a 21%pp-crudo. (figura 3.44)

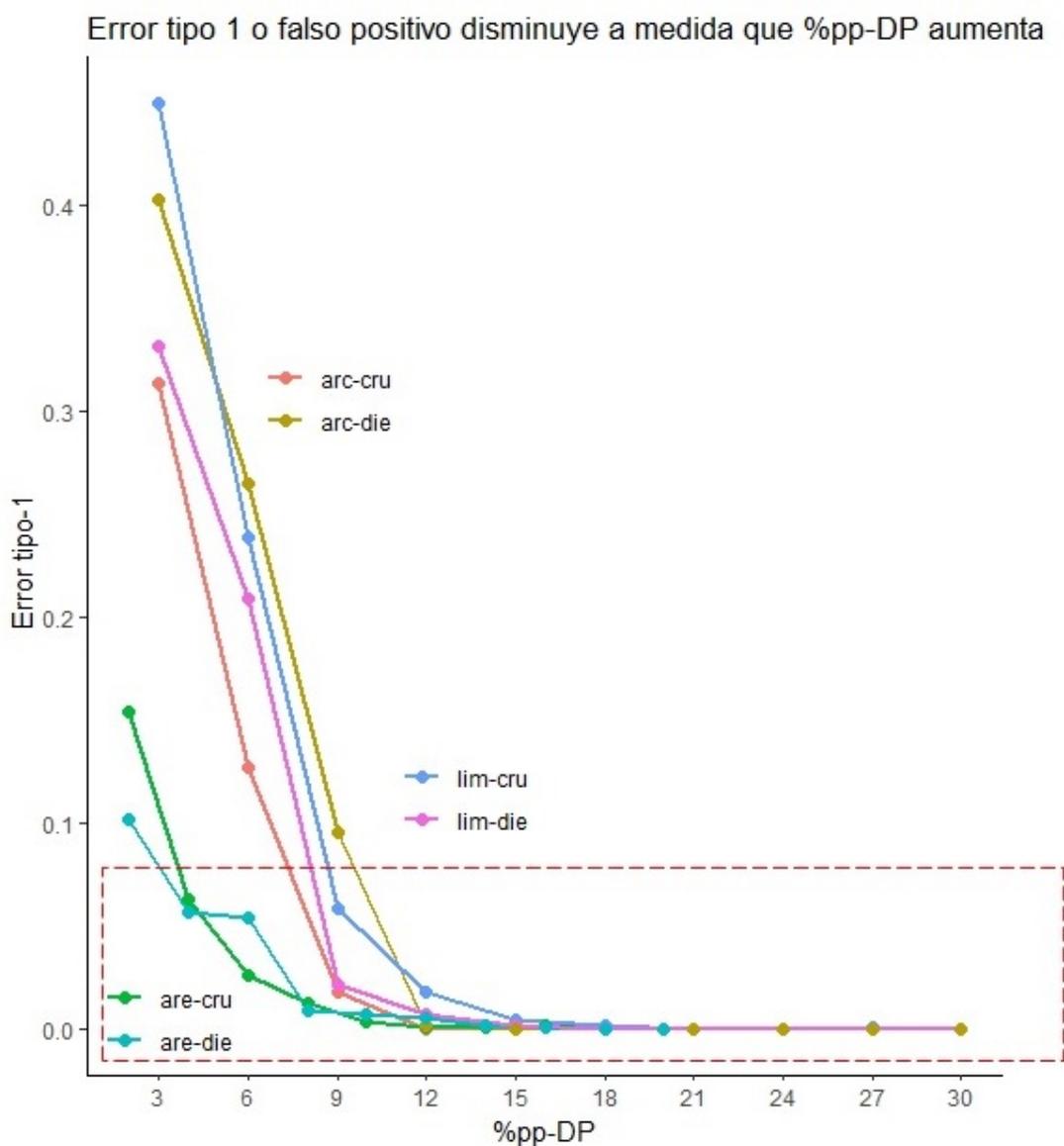


Figura 3.44: Probabilidad de cometer el error tipo 1 cuando las concentraciones del suelo con contaminante cambian.

- Suelo limoso-diésel: (probabilidad del error tipo 1 o tipo 2) < 5% se da para concentraciones mayores a 9%pp-DP y estas probabilidades son nulas para valores de concentración igual o mayores a 18%pp-diésel. (figura 3.44)
- suelo arcilloso-crudo: (probabilidad del error tipo 1 o tipo 2) < 5% se da para concentraciones mayores a 12%pp-DP y estas probabilidades son nulas para valores de concentración igual o mayores a 15%pp-crudo (figura 3.44).
- Suelo arcilloso-diésel: (probabilidad del error tipo 1 o tipo 2) < 5% se da para concentraciones mayores a 12%pp-DP y estas probabilidades son nulas para valores de concentración igual o mayores a 15%pp-diésel (figura 3.44).

- Suelo arenoso-crudo: (probabilidad del error tipo 1 o tipo 2) < 5% se da para concentraciones mayores a 6%pp-DP y estas probabilidades son nulas para valores de concentración igual o mayores a 8%pp-crudo (figura 3.44).
- Suelo arenoso-diésel: (probabilidad del error tipo 1 o tipo 2) < 5% se da para concentraciones mayores a 8%pp-DP y estas probabilidades son nulas para valores de concentración igual o mayores a 10%pp-diésel (figura 3.44).

Tomando como ejemplo el mosaico 1 para poder visualizar por medio de imágenes como se diferenciaría los niveles de concentraciones respecto a un suelo limoso no contaminado, observamos en la figura 3.45 la cantidad de píxeles a los cuales se puede clasificar los suelos no contaminados con los suelos contaminados, usando los niveles críticos correspondientes a la tabla 3.7 y las probabilidades de errores se llega a la siguiente conclusión:

para evitar los errores mencionados se debe clasificar por medio del nivel crítico, esto hace que los píxeles de los PR y blanco disminuyan; pero existe una mejor clasificación entre ellos y a medida que las concentraciones aumentan se clasifican mayor cantidad de píxeles con una menor probabilidad de error.

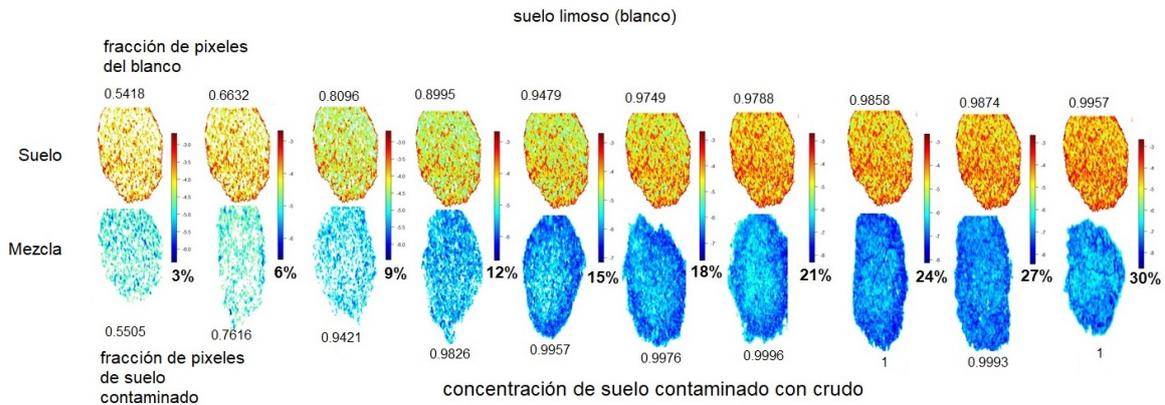


Figura 3.45: Clasificación de las muestras artificiales de suelo limoso con crudo de petróleo con 11 niveles de concentración por medio de los scores de la primera componente.

3.5.4. Análisis de la banda de absorción a 1200 nm en función de la concentración del contaminante

Los suelos no contaminados presentan una banda característica, así como los derivados de petróleo, es de mucha importancia conocer estas regiones antes de poder encontrar una región de absorción de las muestras contaminadas. Los espectros puros de reflectancia y reflectancia-SNV de los suelos limoso, arcilloso, arenoso, del crudo de petróleo y diésel se observan en la figura 3.46.

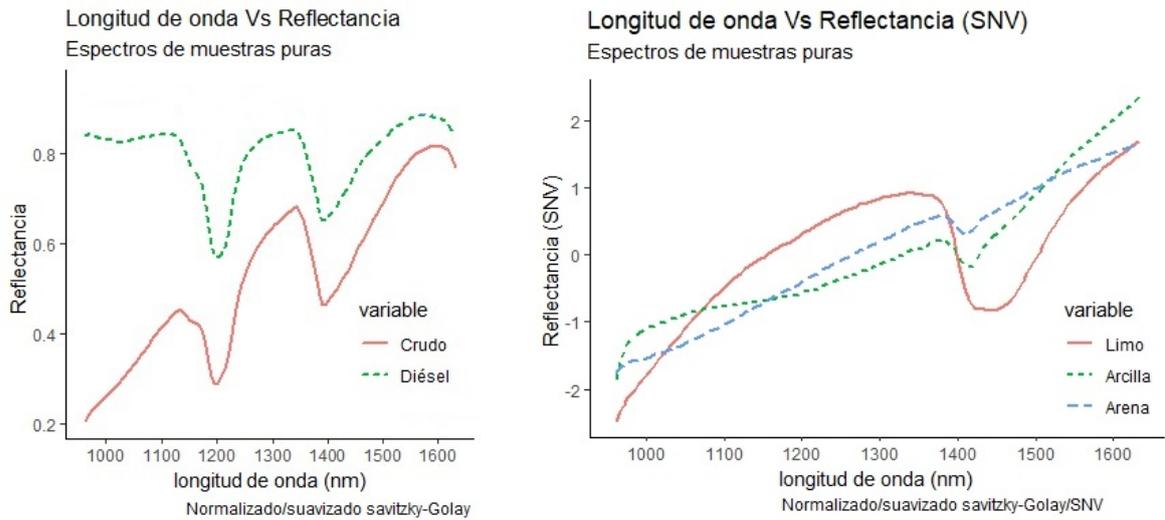


Figura 3.46: (IZQUIERDA) Espectros de reflectancia de los derivados del petróleo, (DERECHA) espectros de suelos.

- En la figura 3.46 observamos los espectros medios de reflectancia del crudo y diésel. Donde el espectro de color rojo del crudo de petróleo es el que se diferencia del diésel en la región de absorción más pronunciada en las bandas próximas a 1189 nm y una inclinación entre la región que se analiza de 961 a 1632 nm (tabla 3.9).
- Se observa que el espectro medio de suelo limoso tiene una región de absorción entre las longitudes de onda de 1418 nm más angosta y pronunciada que los suelos arcillosos y arenoso en la misma región, esto se resume en la tabla 3.9.

Tabla 3.9: Espectros característicos de las muestras

Tipo	Muestras	Longitudes de onda (nm)
Suelos	Limoso, arcilloso, arenoso	1418 corresponde al H ₂ O
Derivados de petróleo	Crudo ligero Diésel	1189 sobretono C-H 1388 sobretono C-H

Para poder analizar la relación entre la concentración de derivado de petróleo y la segunda derivada en la reflectancia en una región de interés se analiza los espectros medios obtenidos del preprocesamiento Nor-max/Sv-G(5,47,2)/SNV-D/2D entre las longitudes 1087 a 1374 nm. Los grupos que se analizarán son seis (6) los cuales son: Limoso-crudo, limoso-diésel, arcilloso-crudo, arcilloso-diésel, arenoso-crudo y arenoso-diésel. Se discute a continuación los resultados obtenido para cada grupo.

- Suelo limoso con crudo de petróleo: En la figura 3.47 se observa los espectros de reflectancia SNV-D entre las longitudes de onda de 1087 a 1374 nm en ella presentan una región de absorción en la longitud de onda de 1205 nm, en esta longitud de onda se observa que se forma una zona de absorción máxima a medida que el nivel de concentración aumenta. La segunda derivada de los espectros de reflectancia SNV-D se visualizan en la figura 3.48 en la cual solo tomamos las longitudes de onda entre 1160 a 1249 nm, estas longitudes de onda nos permitirán encontrar una relación entre los niveles de concentración y el cambio que se dio en el espectro de reflectancia, todos los valores del eje Y en la longitud de 1200 nm (eje X) se utilizaron para poder relacionar con los diferentes niveles de concentración. Al realizar una correlación entre los valores mencionados se observa en la figura 3.49 que los valores se ajustan con una regresión lineal con un coeficiente de determinación $r^2 = 92.57 \%$, lo cual indica que usando un sistema no invasivo podemos encontrar una relación lineal entre la concentración de un contaminante y la reflectancia de la longitud de 1205 nm.

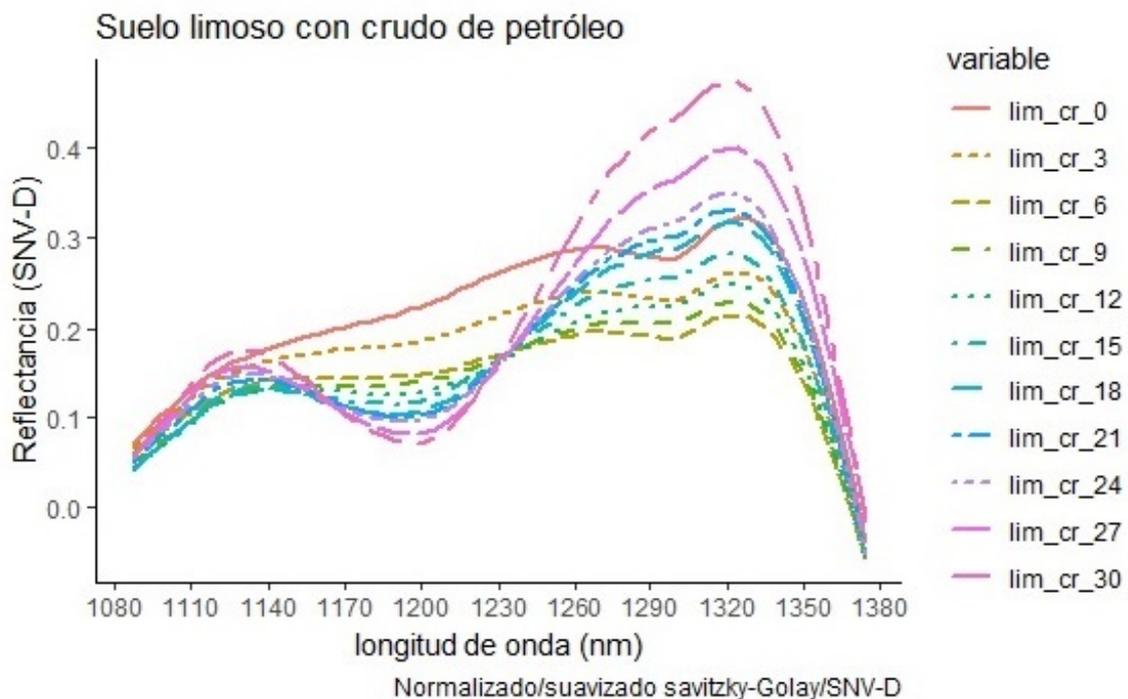


Figura 3.47: Espectro de reflectancia SNV-D del suelo limoso a diferentes niveles de concentración con crudo de petróleo entre la región 1087 a 1374 nm.

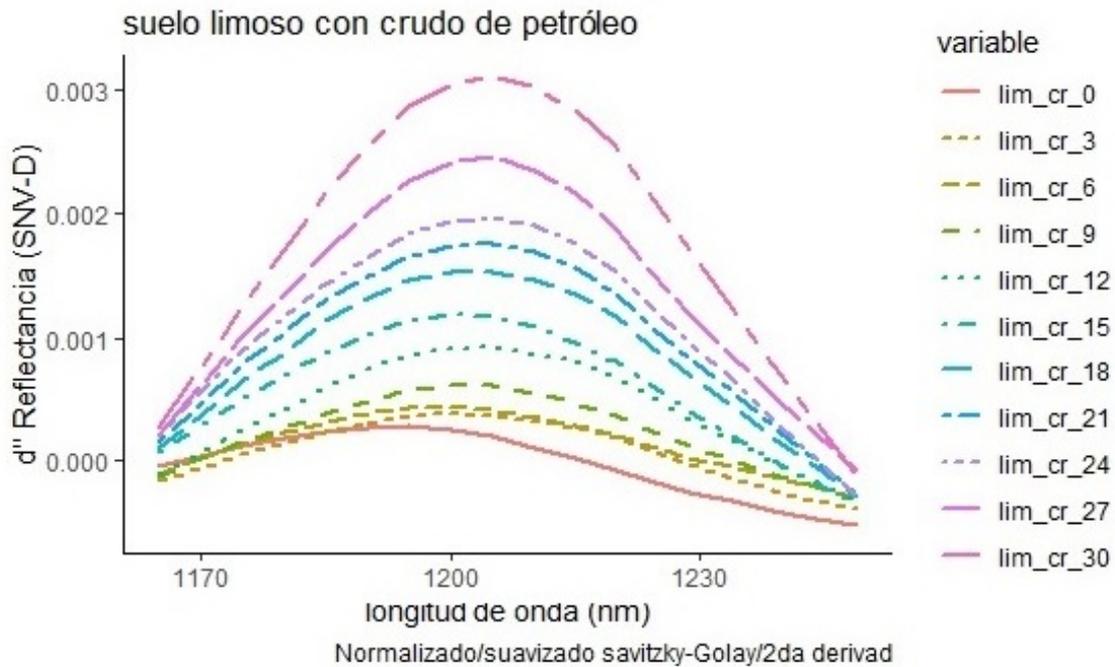


Figura 3.48: Segunda derivada de SNV-D del suelo limoso a diferentes niveles de concentración con crudo de petróleo entre la región 1160 a 1249 nm.

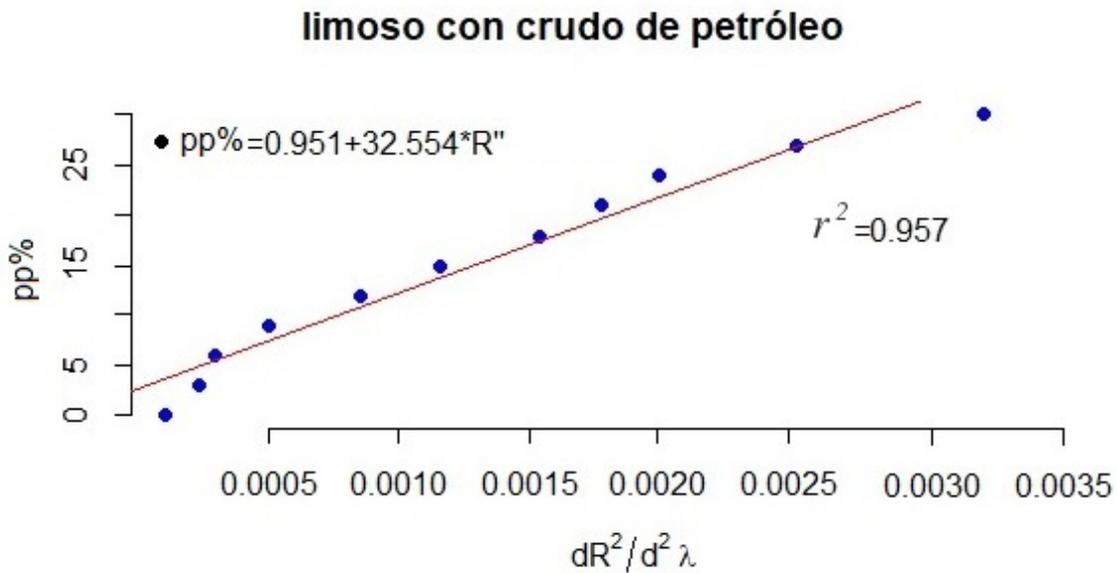


Figura 3.49: Regresión lineal entre las concentraciones y la segunda derivada de la reflectancia SNV-D del suelo limoso con crudo de petróleo en la longitud de onda 1205 nm.

- Suelo limoso con diésel: En la figura 3.50 se muestra los espectros de reflectancia SNV-D entre las longitudes de 1087 a 1374 nm en ella presentan una región de

absorción en la longitud de onda de 1200 nm, en esta longitud de onda se observa que se forma una zona de absorción máxima a medida que el nivel de concentración aumenta. La segunda derivada de los espectros de reflectancia SNV-D se visualizan en la figura 3.51 en la cual solo tomamos la longitudes de onda entre 1126 a 1264 nm, estas curvas nos permitirá encontrar una relación entre los niveles de concentración y el cambio que se dio en el espectro de reflectancia, todos los valores del eje Y en la longitud de onda de 1200 nm (eje X) se utilizaron para poder relacionar con los diferentes niveles de concentración. Al realizar una correlación entre los valores mencionados se observa en la figura 3.52 que los valores se ajustan con una regresión lineal con un coeficiente de determinación $r^2 = 99.20\%$, lo cual indica que usando un sistema no invasivo podemos encontrar una relación lineal entre la concentración de un contaminante y la reflectancia de la longitud de onda de 1200 nm.

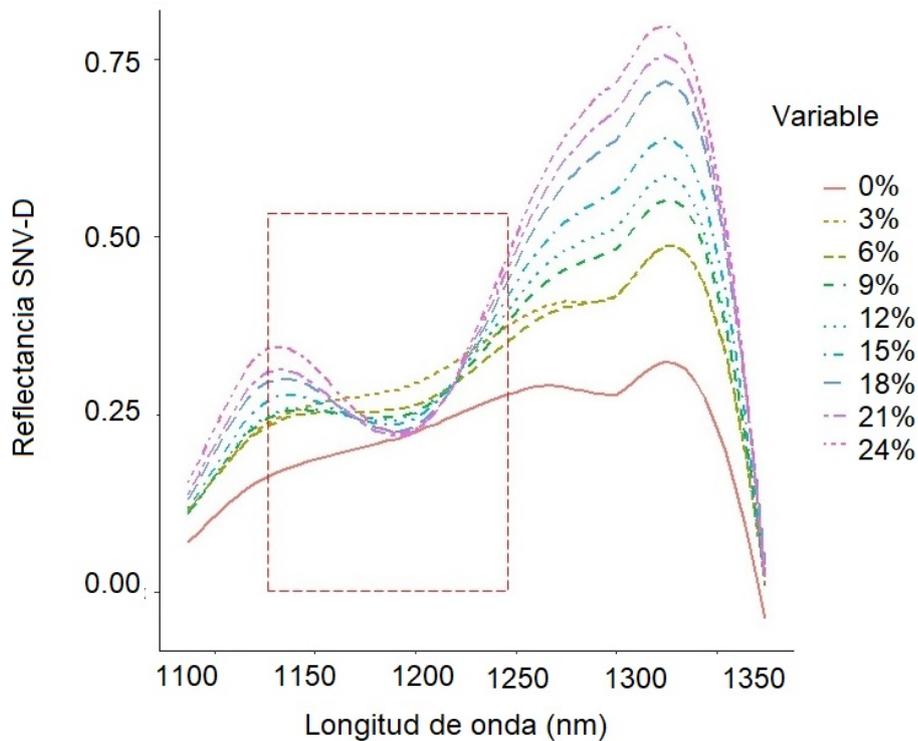


Figura 3.50: Espectro medio de reflectancia SNV-D del suelo limoso con diésel a diferentes niveles de concentración entre la región 1087 a 1374 nm.

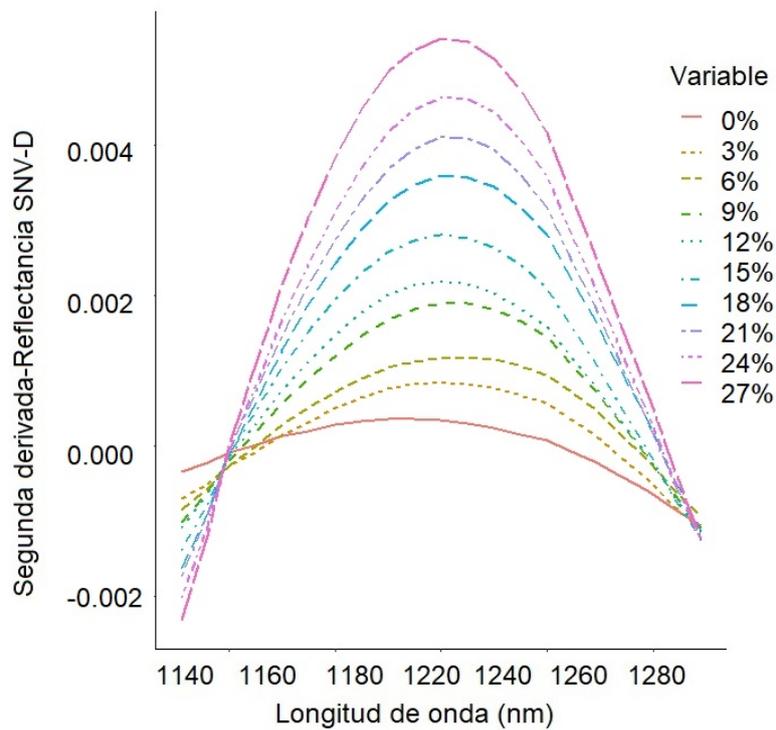


Figura 3.51: Segunda derivada de SNV-D del suelo limoso a diferentes niveles de concentración con diésel entre la región 1126 y 1264 nm.

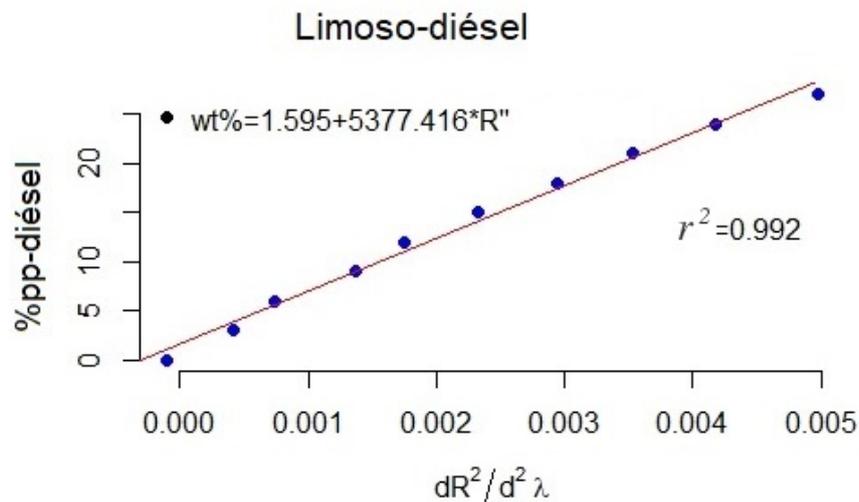


Figura 3.52: Regresión lineal entre las concentraciones y la segunda derivada de la reflectancia SNV-D del suelo limoso con crudo de petróleo en la longitud de onda 1200 nm.

- Suelo arcilloso con crudo de petróleo y diésel: En la figura 3.53 y figura 3.54 se muestra los espectros de reflectancia SNV-D entre las longitudes de 1087 a 1374 nm para el crudo y diésel en ella presentan una región de absorción en la longitud de onda de 1224 y 1200 nm respectivamente, en esta longitud de onda

se observa que se forma una zona de absorción máxima a medida que el nivel de concentración aumenta. La segunda derivada de los espectros de reflectancia SNV-D se visualizan en la figura 3.55 y figura 3.56 para el crudo y diésel a las cuales tomamos las longitudes de onda entre 1121 a 1259 nm y 1126 a 1279 nm respectivamente, estas curvas nos permite encontrar una relación entre los niveles de concentración y el cambio que se dio en el espectro de reflectancia, los valores del eje Y en la longitud de 1224 y 1200 nm (eje X) para el crudo y diésel respectivamente se utilizaron para relacionarlos con los diferentes niveles de concentración respectivamente. Al realizar una correlación entre los valores mencionados, se observa en la figura 3.57 y figura 3.58 que los valores se ajustan con una regresión lineal con un coeficiente de determinación $r^2 = 93.10\%$ y $r^2 = 90.00\%$, lo cual indica que usando un sistema de imágenes hiperespectrales no invasivo podemos encontrar una relación lineal entre la concentración del crudo o diésel en el suelo arcilloso y la segunda derivada de la reflectancia de la longitud de onda en 1200 nm.

- Algunos detalles que se encuentran en los suelos arcillosos con crudo son: Los espectros que tiene concentración mayor a 18% llegan a superponerse y las regiones de absorción entre 1121 a 1259 nm y 1126 a 1279 nm para el crudo y diésel no aumenta a medida que la concentración aumenta (figura 3.53 figura 3.54).

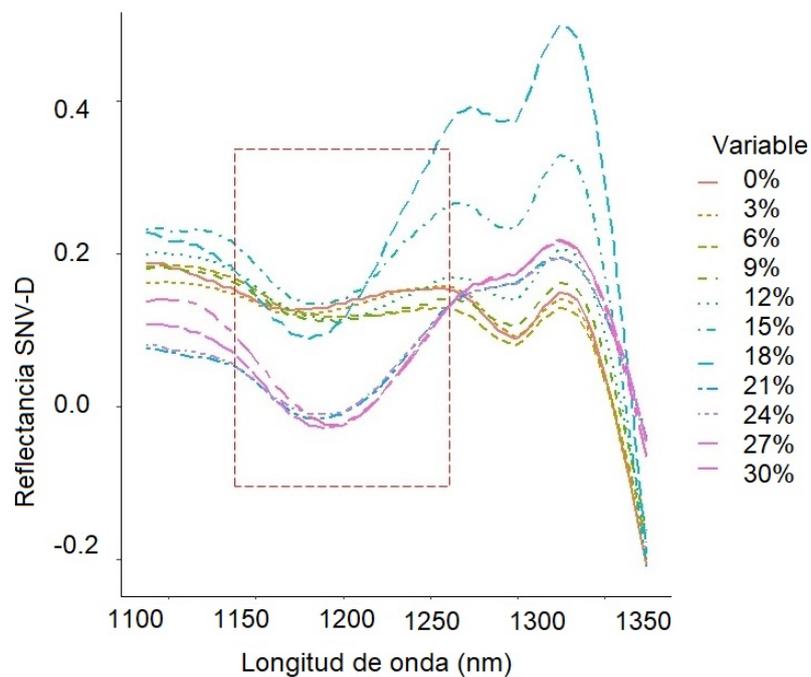


Figura 3.53: Espectro de reflectancia SNV-D del suelo arcilloso a diferentes niveles de concentración con crudo de petróleo entre la región 1087 a 1374 nm.

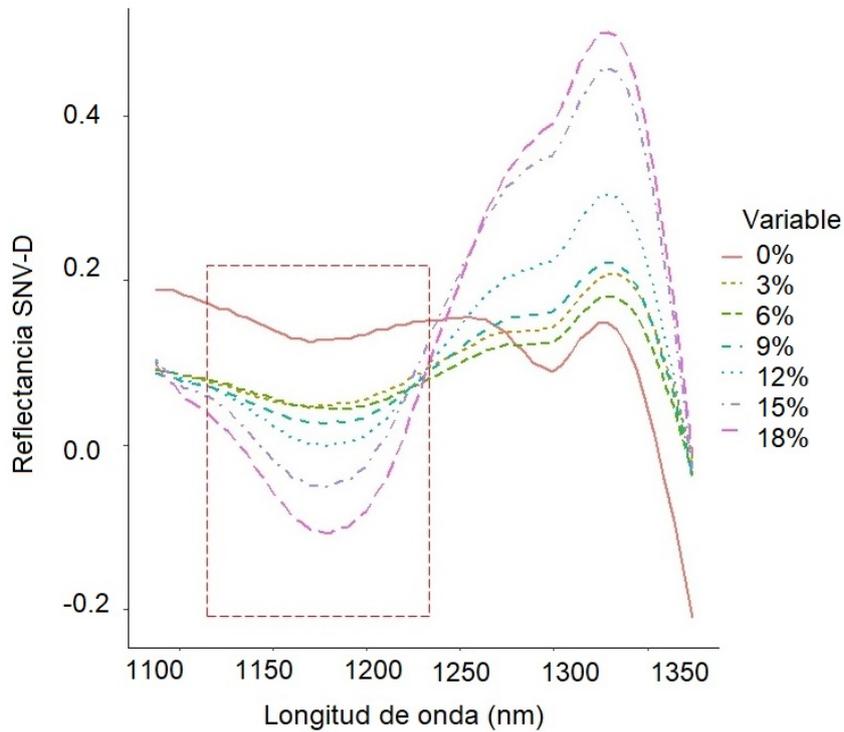


Figura 3.54: Espectro de reflectancia SNV-D del suelo arcilloso a diferentes niveles de concentración con diésel entre la región 1087 a 1374 nm.

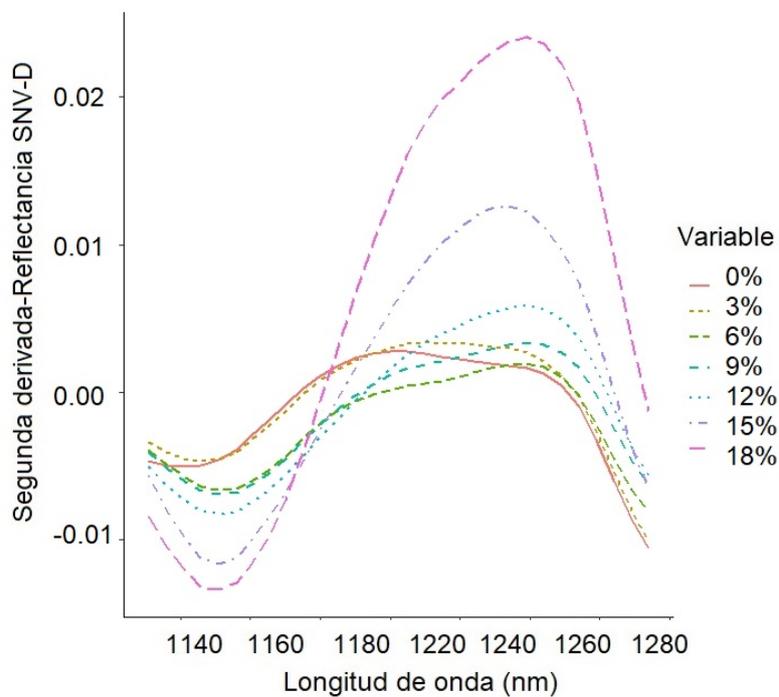


Figura 3.55: Segunda derivada de SNV-D del suelo arcilloso a diferentes niveles de concentración con diésel entre la región 1121 a 1259 nm.

- Solo para el suelo arcilloso contaminado con crudo y diésel se toma las concen-

tración que son igual y menor a 18 %, partiendo del hecho que las muestras se saturan de los derivados de petróleo a valores mayores a 18 % (figura 3.55 y figura 3.56).

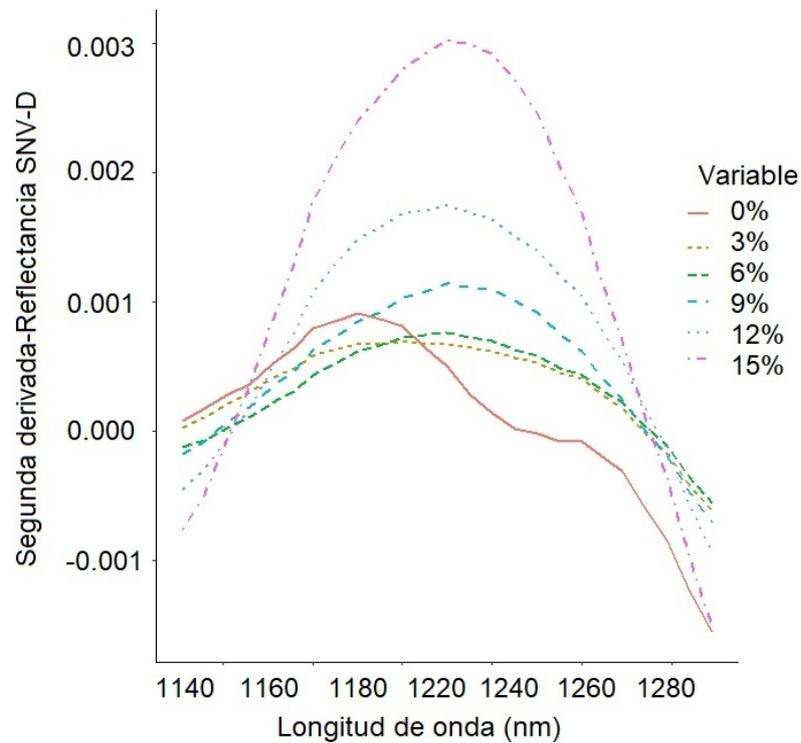


Figura 3.56: Segunda derivada de SNV-D del suelo arcilloso a diferentes niveles de concentración con diésel entre la región 1126 a 1279 nm.

- Se realizaron las correlaciones con los 11 valores (incluye el suelo no contaminado) de concentración para el crudo y diésel de forma separada con los respectivos valores de la segunda derivada de reflectancia, al realizarlos el coeficiente de correlación estaba por debajo del 80 %, por lo tanto se excluyeron los valores mayores a 18 % de la correlación.

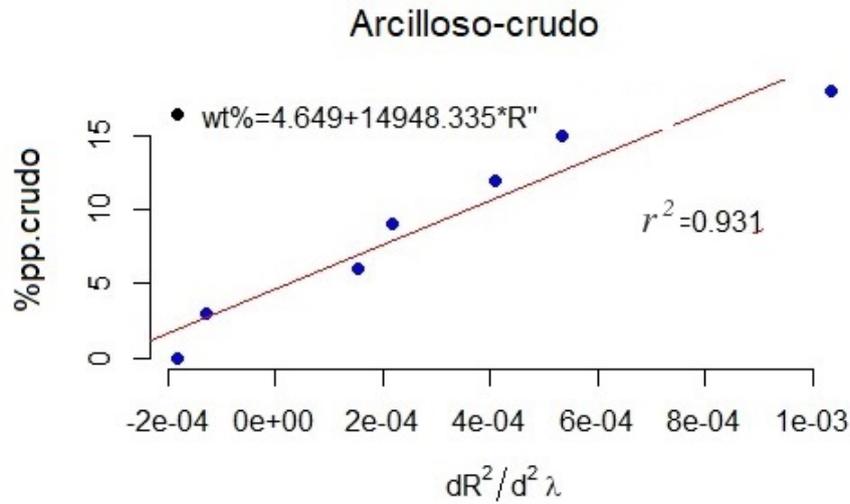


Figura 3.57: Regresión lineal entre la concentraciones y la segunda derivada de la reflectancia SNV-D del suelo arcilloso con crudo de petróleo en la longitud de onda 1224 nm.

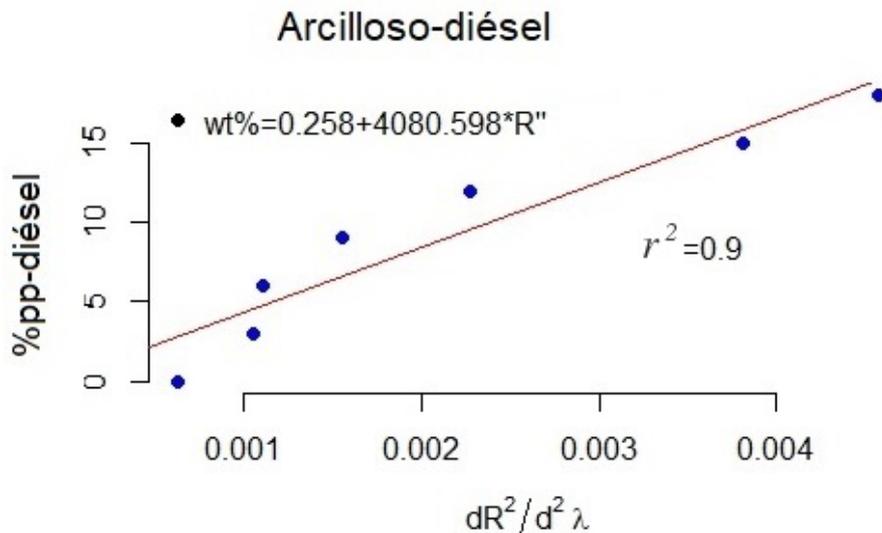


Figura 3.58: Regresión lineal entre la concentraciones y la segunda derivada de la reflectancia SNV-D del suelo arcilloso con diésel en la longitud de onda 1200 nm.

- Suelo arenoso con crudo de petróleo y diésel: En la figura 3.59 y figura 3.60 se muestra los espectros de reflectancia SNV-D entre las longitudes de 1087 a 1374 nm para el crudo y diésel en ella presentan una región de absorción en la longitud de onda de 1200 nm para ambos casos. En esta longitud de onda se observa que se forma una zona de absorción máxima a medida que el nivel de concentración aumenta. La segunda derivada de los espectros de reflectancia SNV-D se visualizan en la figura 3.61 y figura 3.62 para el crudo y diésel a las cuales tomamos

las longitudes de onda entre 1131 a 1254 nm y 1126 a 1259 nm respectivamente. Estas curvas nos permiten encontrar una relación entre los niveles de concentración y el cambio que se dio en el espectro de reflectancia, los valores del eje Y en la longitud de 1200 nm (eje X) para el crudo y diésel se utilizaron para relacionarlos con los diferentes niveles de concentración respectivamente. Al realizar una correlación entre los valores mencionados, se observa en la figura 3.63 y figura 3.64 que los valores se ajustan con una regresión lineal con coeficientes de determinación de $r^2 = 99.31 \%$ y $r^2 = 96.37 \%$, lo cual indica que usando un sistema de imágenes hiperespectrales no invasivo podemos encontrar una relación lineal entre la concentración del crudo o diésel en el suelo arenoso y la segunda derivada de la reflectancia de la longitud de onda en 1200 nm.

- Los espectros entre las longitudes de 1087 a 1374 nm (figura 3.59 y figura 3.60) no presentan una absorción tan pronunciada con respecto a los suelos limosos contaminados; pero de igual manera que los suelos contaminados cumplen una relación lineal entre la segunda derivada de la reflectancia en la longitud de onda de 1200 nm y sus diferentes concentraciones (figura 3.63 y figura 3.64).

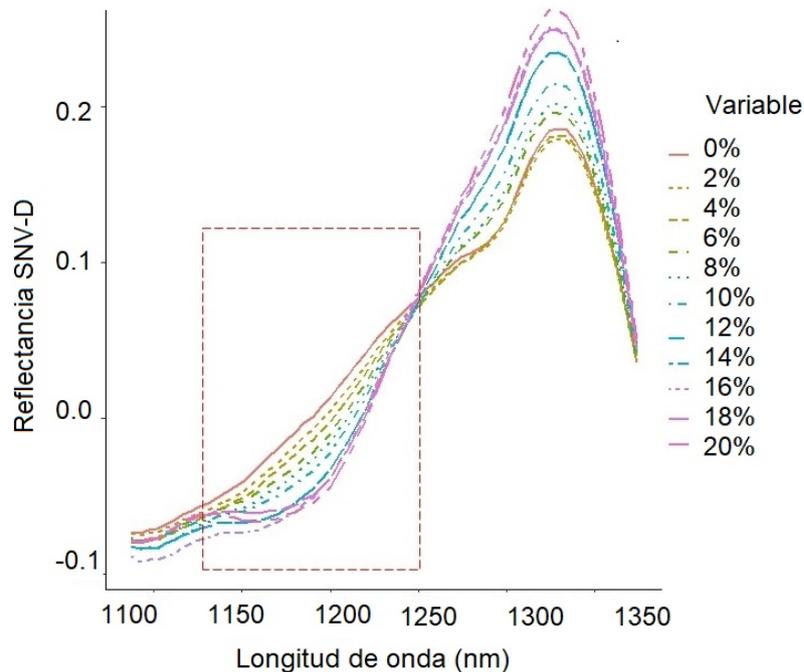


Figura 3.59: Espectro de reflectancia SNV-D del suelo arenoso a diferentes niveles de concentración con crudo de petróleo entre la región 1087 a 1374 nm.

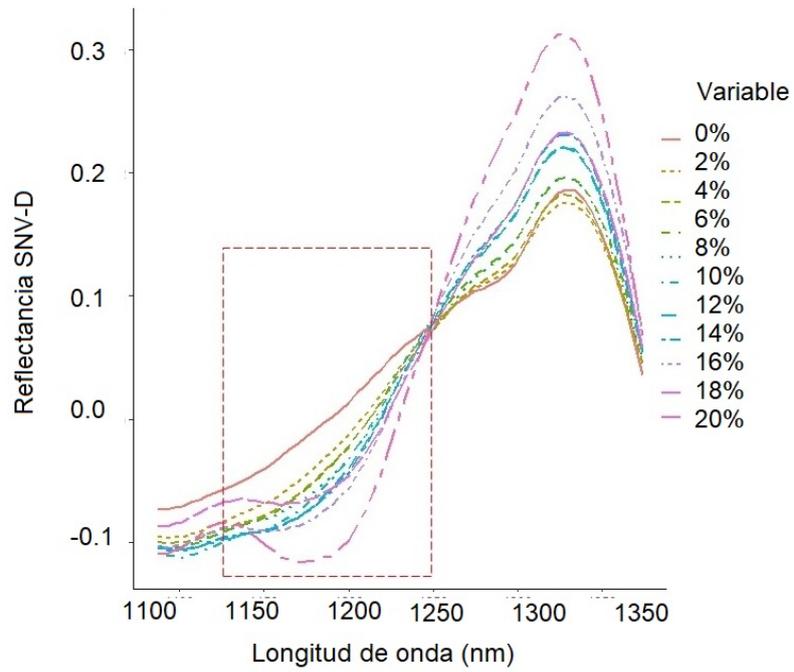


Figura 3.60: Espectro de reflectancia SNV-D del suelo arenoso a diferentes niveles de concentración con diésel entre la región 1087 a 1374 nm.

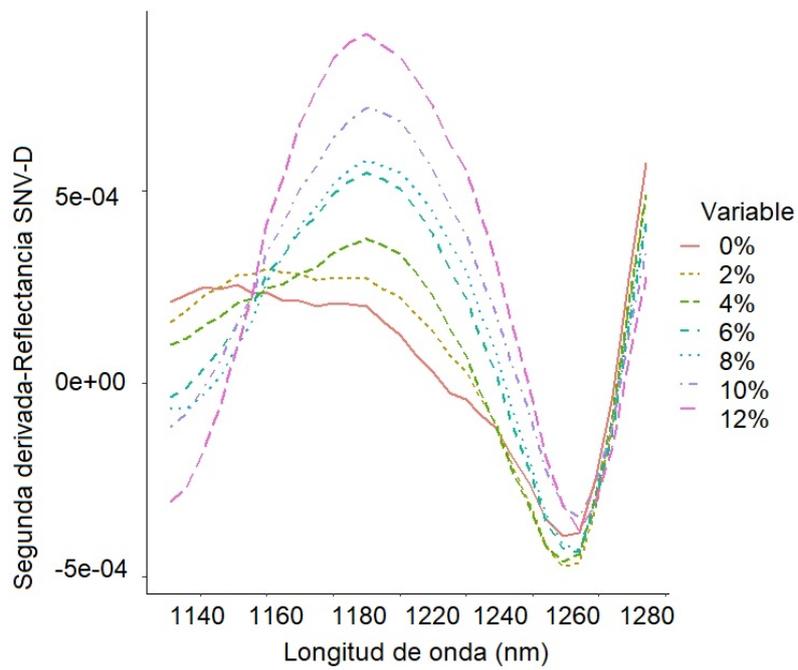


Figura 3.61: Espectro de reflectancia de la segunda derivada SNV-D del suelo arenoso a diferentes niveles de concentración con crudo de petróleo entre la región 1131 y 1254 nm.

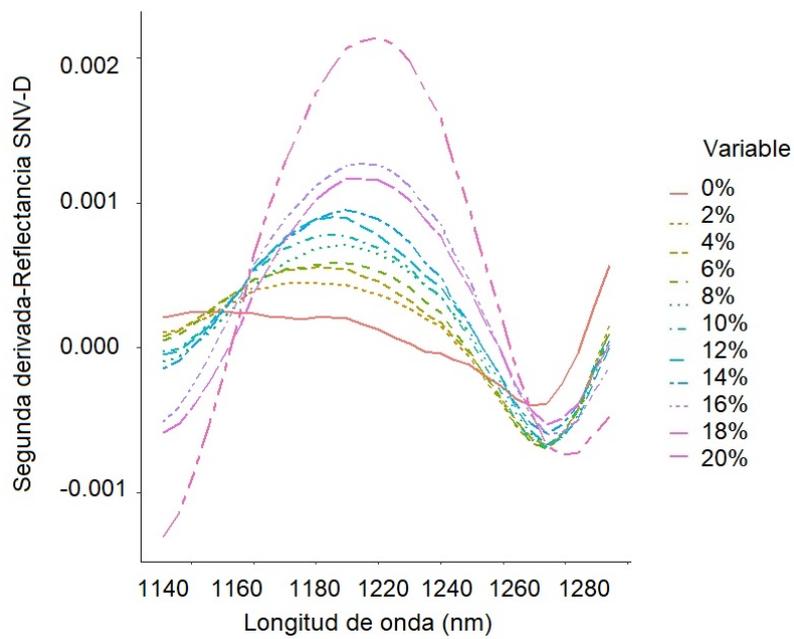


Figura 3.62: Espectro de reflectancia de la segunda derivada SNV-D del suelo arenoso a diferentes niveles de concentración con diésel entre la región 1126 y 1259 nm.

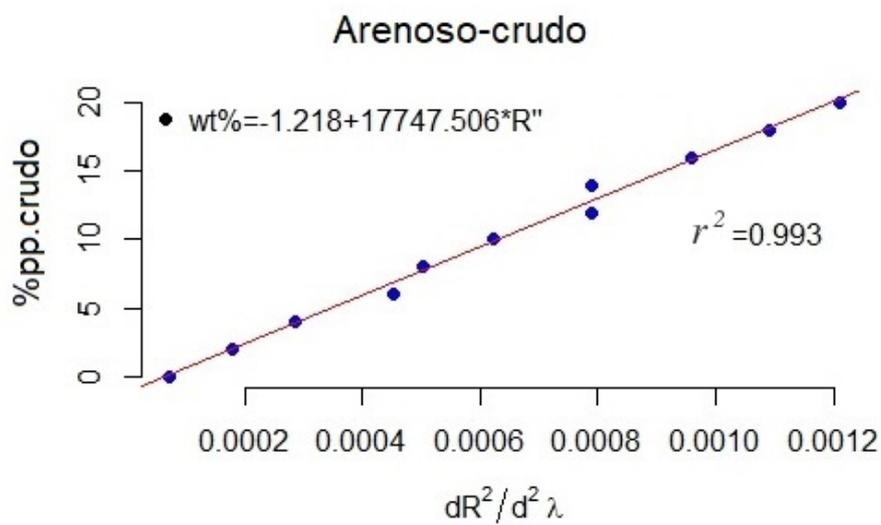


Figura 3.63: Regresión lineal entre las concentraciones y la segunda derivada de la reflectancia SNV-D del suelo arenoso con crudo de petróleo en la longitud de onda 1200 nm.

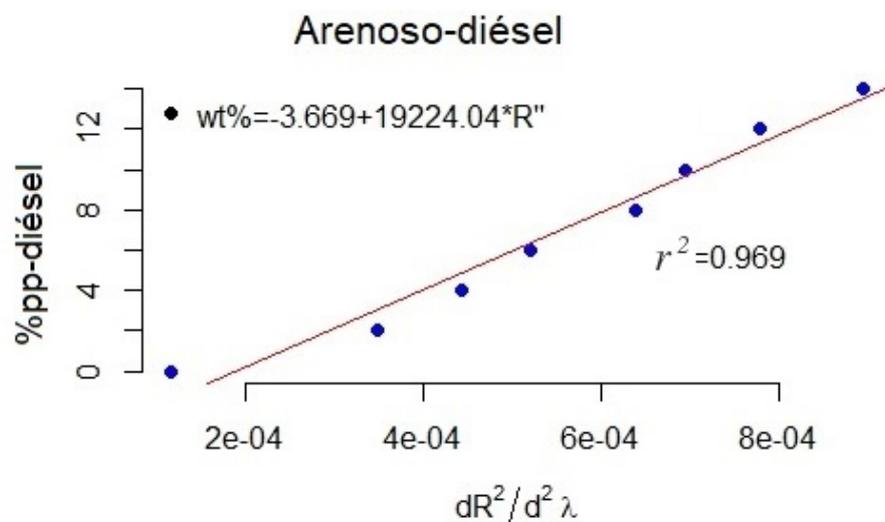


Figura 3.64: Regresión lineal entre las concentraciones y la segunda derivada de la reflectancia SNV-D del suelo arenoso con diésel en la longitud de onda 1200 nm.

Los valores de coeficiente de determinación (r^2) se muestra en la tabla 3.10 para cada tipo de suelo contaminado con crudo de petróleo y diésel con sus respectivas longitudes de onda. Un coeficiente de determinación cercano a 1 nos indica una alta correlación entre la concentración y la segunda derivada SNV-D. Observando que para todos los suelos contaminados existe una relación lineal entre el nivel de concentración y la segunda derivada de la reflectancia en su correspondiente banda.

Tabla 3.10: Grupos con sus respectivas características para el modelo lineal

Grupo	Región (nm)	Longitud (nm)	r^2
lim-cru	1160-1249	1205	0.9257
arc-cru	1121-1259	1224	0.9310
are-cru	1131-1254	1200	0.9931
lim-die	1126-1264	1200	0.9920
arc-die	1126-1279	1200	0.9000
are-die	1126-1259	1200	0.9637

CONCLUSIONES

- En este trabajo, se ha logrado clasificar y cuantificar con éxito suelos limosos, arcillosos y arenosos contaminados con diferentes concentraciones de derivados de petróleo usando imágenes hiperespectrales y aplicando métodos multivariantes. Los resultados muestran que un sistema de imágenes hiperespectrales en el laboratorio, en la región NIR permite clasificar y cuantificar suelos contaminados con respecto a los suelos no contaminados de manera rápida y no destructiva, así como relacionar las diferentes concentraciones de contaminante con la reflectancia de la muestra registrada a 1200 nm.
- Se elaboró por primera vez una biblioteca de imágenes hiperespectrales de suelos limosos, arcillosos y arenosos contaminados con diferentes concentraciones de derivado de petróleo (crudo de petróleo ligero y diésel) obtenidas en el laboratorio. Las concentraciones que se utilizaron están en el rango de 3 % a 30 % para el crudo de petróleo y 2 % a 20 % para el diésel.
- De los resultados obtenidos se tiene que un suelo limoso con concentraciones superiores a 21 % pp-crudo pueden ser clasificados sin ningún tipo de error como suelo contaminado, para el suelo limoso con diésel el valor debe ser superior a 18 %, para el suelo arcilloso con crudo 12 %, para el suelo arcilloso con diésel 15 %, para el suelo arenoso con crudo 6 % y para el suelo arenoso con diésel 8 %. Por lo tanto, los suelos arenosos contaminados con cualquiera de los dos derivados de petróleo son los que se clasifican mejor a pequeñas concentraciones de contaminante de derivado de petróleo.
- El sistema de imágenes hiperespectrales identificó los cambios de algunos píxeles en la mínima concentración de 3 % de contaminante en el suelo limoso y arcilloso, de la misma manera para el suelo arenoso con mínima concentración de 2 %, esto indica que a concentraciones bajas se puede visualizar y ver cambios cuantitativos entre el suelo no contaminado y contaminado.
- Existe relación lineal entre las concentraciones y la segunda derivada de la reflectancia SNV-D para longitudes de onda de máxima absorción entre la región de 1160 y 1249 nm. Los coeficientes de determinación son mayores a 90.00 % y sólo para los suelos arcillosos contaminados con los DP se han tomado hasta la concentración de 18 %pp-DP.
- El método de imágenes hiperespectrales tiene un costo operativo relativamente bajo en comparación con los métodos tradicionales y proporciona una evaluación

rápida del grado de contaminación de un suelo. Los resultados positivos obtenidos con experimentos controlados también abren la oportunidad de explorar las capacidades de esta herramienta en áreas a campo abierto, así como el monitoreo de zonas afectadas. Es de importancia realizar más investigaciones cualitativas y cuantitativas sobre la correlación entre el contenido de petróleo en un suelo y los datos espectrales. Los hallazgos en este estudio pueden considerarse como referenciales o línea de base para futuras aplicaciones usando imágenes hiperespectrales en relación a los contaminantes en suelos.

- Los trabajos futuros son: la validación del uso de imágenes hiperespectrales para suelos contaminados, mediciones en una UAV para clasificar suelos no contaminados en la superficie del suelo y uso de algoritmo de IA para una clasificación supervisada.

BIBLIOGRAFIA

- AIE (2019). Demanda diaria de petróleo crudo a nivel mundial de 2006 a 2019. 18
- Amigo, J. M., Babamoradi, H., and Elcoroaristizabal, S. (2015). Hyperspectral image analysis. a tutorial. *Analytica chimica acta*, 896:34–51. 16, 22, 39
- Analytik (2019). <https://analytik.co.uk/what-is-remote-sensing/>. *analytik.co.uk*, page 01 de Julio. 6, 26
- Andreoli, G., Bulgarelli, B., Hosgood, B., and Tarchi, D. (2007). Hyperspectral analysis of oil and oil-impacted soils for remote sensing purposes. *European Commission Joint Research Centre: Luxembourg*, page 36. 16, 20
- Asadzadeh, S. and de Souza Filho, C. R. (2016). Investigating the capability of worldview-3 superspectral data for direct hydrocarbon detection. *Remote sensing of environment*, 173:162–173. 19
- Bajorski, P. (2011). *Statistics for imaging, optics, and photonics*, volume 808. John Wiley & Sons. 6, 21, 23, 24, 32
- Barnes, R., Dhanoa, M. S., and Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied spectroscopy*, 43(5):772–777. 39
- Botello, A. V. (2005). Características composición y propiedades fisicoquímicas del petróleo. *Golfo de México Contaminación e Impacto Ambiental: Diagnóstico y Tendencias*, 1(1):261–268. 48, 49
- Cao, F., Yang, Z., Ren, J., Jiang, M., and Ling, W.-K. (2017). Does normalization methods play a role for hyperspectral image classification? *arXiv preprint arXiv:1710.02939*. 21, 35, 38
- Castro, W. (2015). *Aplicación de la tecnología de imágenes hiperespectrales al control de calidad de productos agroalimentarios de la región de amazonas (Perú)*. PhD thesis, Universidad Politécnica de Valencia. 6, 28, 29
- Catelli, E., Randeberg, L. L., Strandberg, H., Alsberg, B. K., Maris, A., and Vikki, L. (2018). Can hyperspectral imaging be used to map corrosion products on outdoor bronze sculptures? *NA*. 25

- Chakraborty, S., Weindorf, D. C., Morgan, C. L., Ge, Y., Galbraith, J. M., Li, B., and Kahlon, C. S. (2010). Rapid identification of oil-contaminated soils using visible near-infrared diffuse reflectance spectroscopy. *Journal of environmental quality*, 39(4):1378–1387. 16
- Correal, C. D. and García, J. (2012). Aspectos físicos a considerar en la calibración radiométrica de imágenes satelitales. *UD y la geomática*, NA(6):11–18. 26
- Dai, Q., Cheng, J.-H., Sun, D.-W., and Zeng, X.-A. (2014). Potential of hyperspectral imaging for non-invasive determination of mechanical properties of prawn (*metapenaeus ensis*). *Journal of Food Engineering*, 136:64–72. 16
- El-Rahman, S. A. and Zolait, A. H. S. (2018). Hyperspectral image analysis for oil spill detection: a comparative study. *International Journal of Computing Science and Mathematics*, 9(2):103–121. 16
- Falco, N. (2015). *Advanced Spectral and Spatial Techniques for Hyperspectral Image Analysis and Classification*. PhD thesis, University of Trento. 6, 23, 24
- Fan, R., Xing, L., Pan, J., Shan, X., You, J., Li, C., and Zhong, W. (2019). Study of the relationship between the oil content of oil sands and spectral reflectance based on spectral derivatives. *Journal of the Indian Society of Remote Sensing*, pages 1–10. 16, 37
- Felipó, M. (1992). Contaminación del suelo e impacto ambiental. *Seminario “Contaminación, protección y saneamientos de suelos”*. UIPM, Valencia. 49
- Feng, C.-H., Makino, Y., Yoshimura, M., Thuyet, D. Q., and García-Martín, J. F. (2018). Hyperspectral imaging in tandem with r statistics and image processing for detection and visualization of ph in japanese big sausages under different storage conditions. *Journal of food science*, 83(2):358–366. 24
- García, V. J., Márquez, C. O., Cedeño, A. R., and Montesdeoca, K. G. (2019). Assessing bioremediation of soils polluted with fuel oil 6 by means of diffuse reflectance spectroscopy. *Resources*, 8(1):36. 16
- Gisbert Blanquer, J. M., Ibañez Asensio, S., and Moreno Ramón, H. (2010). La textura del suelo. 6, 45, 46
- Gowen, A., O’Donnell, C., Cullen, P., Downey, G., and Frias, J. (2007). Hyperspectral imaging—an emerging process analytical tool for food quality and safety control. *Trends in food science & technology*, 18(12):590–598. 29

- Guide, S. Q. T. K. (1999). Soil quality test kit guide. 18
- Hecht, E. (2000). *Óptica*. Pearson Educación,. 6, 11, 42, 43
- Kühn, F., Oppermann, K., and Hörig, B. (2004). Hydrocarbon index—an algorithm for hyperspectral detection of hydrocarbons. *International Journal of Remote Sensing*, 25(12):2467–2473. 16
- Lassalle, G., Credoza, A., Hédacq, R., Fabre, S., Dubucq, D., and Elger, A. (2018). Assessing soil contamination due to oil and gas production using vegetation hyperspectral reflectance. *Environmental science & technology*, 52(4):1756–1764. 16
- Li, X., Li, R., Wang, M., Liu, Y., Zhang, B., and Zhou, J. (2018a). Hyperspectral imaging and their applications in the nondestructive quality assessment of fruits and vegetables. *Hyperspectral Imaging in Agriculture, Food and Environment*, page 27. 6, 27
- Li, X., Wei, Y., Xu, J., Feng, X., Wu, F., Zhou, R., Jin, J., Xu, K., Yu, X., and He, Y. (2018b). Ssc and ph for sweet assessment and maturity classification of harvested cherry fruit based on nir hyperspectral imaging technology. *Postharvest Biology and Technology*, 143:112–118. 6, 33
- Long, J., Wang, K., Yang, M., and Zhong, W. (2019). Rapid crude oil analysis using near-infrared reflectance spectroscopy. *Petroleum Science and Technology*, 37(3):354–360. 16
- Lu, G. and Fei, B. (2014). Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1):010901. 16
- Manolakis, D., Marden, D., and Shaw, G. A. (2003). Hyperspectral image processing for automatic target detection applications. *Lincoln laboratory journal*, 14(1):79–116. 6, 21, 41, 47
- Manolakis, D. G., Lockwood, R. B., and Cooley, T. W. (2016). *Hyperspectral imaging remote sensing: physics, sensors, and algorithms*. Cambridge University Press. 6, 28, 29, 30, 39, 41
- Martínez, E. I. (2015). *Desarrollo de técnicas de visión hiperespectral y tridimensional para el sector agroalimentario*. PhD thesis, NA. 6, 11, 16, 23, 50
- Martínez, V. E., López, M., et al. (2001). Efecto de hidrocarburos en las propiedades físicas y químicas de suelo arcilloso. *Terra latinoamericana*, 19(1). 20

- MathWorks (2019). Datos de detendencia. url<https://la.mathworks.com>. Accedido 24/04/2019. 6, 38
- Minagri (2019). Producción de petróleo. *Ministerio de Agricultura y Riego del Perú*, page 1. 18
- Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., and Scheunders, P. (2017). Close range hyperspectral imaging of plants: a review. *Biosystems engineering*, 164:49–67. 6, 24, 25, 26
- Ozaki, Y., McClure, W. F., and Christy, A. A. (2006). *Near-infrared spectroscopy in food science and technology*. John Wiley & Sons. 34
- Park, B. and Lu, R. (2015). *Hyperspectral imaging technology in food and agriculture*. Springer. 16, 21, 22, 25, 32, 37, 39, 41, 44
- Pelta, R. and Ben-Dor, E. (2019). Assessing the detection limit of petroleum hydrocarbon in soils using hyperspectral remote-sensing. *Remote Sensing of Environment*, 224:145–153. 16, 20, 25
- Reséndez-Hernández, L. A., Prudencio-Csapek, D., and Lozano-García, D. F. (2018). Hyperspectral analysis of soil polluted with four types of hydrocarbons. *Geocarto International*, pages 1–18. 16, 18
- Resonon (2017). *SpectrononPro Manual*. Resonon Inc, Bozeman, MT 59715, 5.1 edition. 21, 53
- Savitzky, A. and Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639. 21, 36
- Schwartz, G., Ben-Dor, E., and Eshel, G. (2013). Quantitative assessment of hydrocarbon contamination in soil using reflectance spectroscopy: a “multipath” approach. *Applied spectroscopy*, 67(11):1323–1331. 16
- Serway, R. A. and Jewett, J. W. (2018). *Physics for scientists and engineers with modern physics*. Cengage learning. 42
- Siesler, H. W., Ozaki, Y., Kawata, S., and Heise, H. M. (2008). *Near-infrared spectroscopy: principles, instruments, applications*. John Wiley & Sons. 51
- Solano, H. L. and Álvarez, C. R. (2005). *Estadística descriptiva y distribuciones de probabilidad*. Universidad del Norte. 72

- Wang, W. and Peng, Y. (2018). Hyperspectral imaging for assessing quality and safety of meat. In *Hyperspectral Imaging in Agriculture, Food and Environment*. IntechOpen. 6, 22
- Williams, P., Norris, K., et al. (1987). *Near-infrared technology in the agricultural and food industries*. American Association of Cereal Chemists, Inc. 34
- Winkelmann, K. H. (2005). On the applicability of imaging spectrometry for the detection and investigation of contaminated sites with particular consideration given to the detection of fuel hydrocarbon contaminants in soil. *NA, NA(NA):NA*. 6, 47, 48, 51
- Yin, W., Zhang, C., Zhu, H., Zhao, Y., and He, Y. (2017). Application of near-infrared hyperspectral imaging to discriminate different geographical origins of chinese wolfberries. *PloS one*, 12(7):e0180534. 24
- Yoon, S.-C., Windham, W. R., Ladely, S. R., Heitschmidt, J. W., Lawrence, K. C., Park, B., Narang, N., and Cray, W. C. (2013). Hyperspectral imaging for differentiating colonies of non-0157 shiga-toxin producing escherichia coli (stec) serogroups on spread plates of pure cultures. *Journal of Near Infrared Spectroscopy*, 21(2):81–95. 32
- Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., and Liu, C. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Research International*, 62:326–343. 29, 30
- Zhu, H., Chu, B., Fan, Y., Tao, X., Yin, W., and He, Y. (2017). Hyperspectral imaging for predicting the internal quality of kiwifruits based on variable selection algorithms and chemometric models. *Scientific reports*, 7(1):7845. 24

Anexo A: Manejo del software SpectronPro

Manejo del software SpectronPro

El software Spectron Pro está disponible como descarga gratuita. Permite la adquisición de datos de los sistemas de imágenes hiperespectrales del equipo.

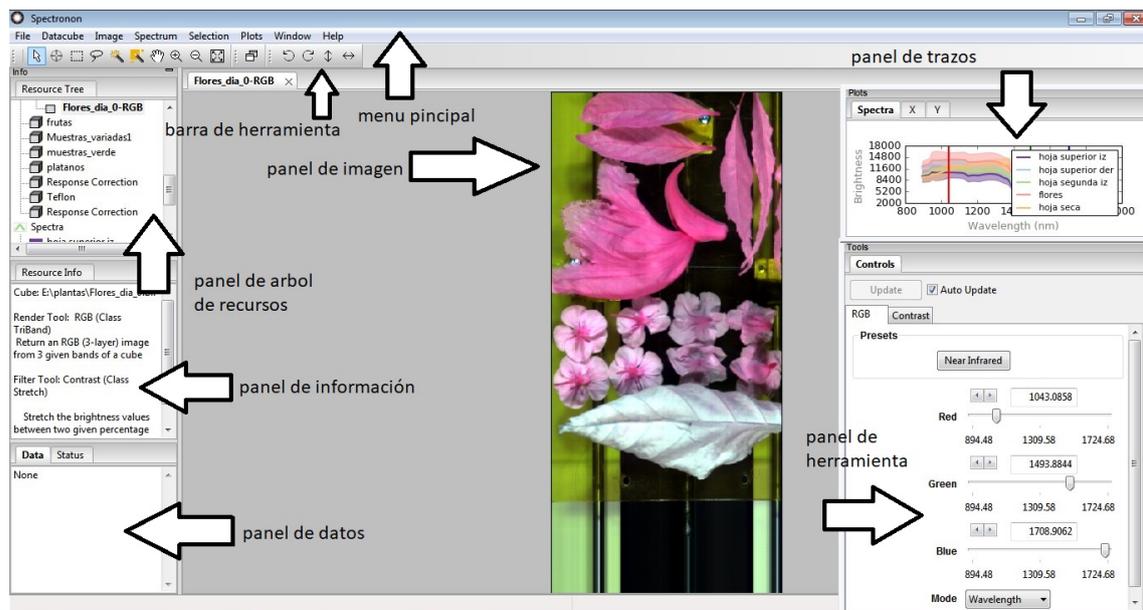


Figura 3.65: Software Spectron Pro.

- **Menú principal:** En el cual se puede encontrar las opciones habituales de cualquier software, sea las de guardar el archivo, exportar, importar, abrir un data cube, guardar las imágenes, guardar los espectros y diversas opciones relacionadas al entorno del imágenes y espectros, las cuales permite llevar a otro entorno para cualquier propósito de análisis.
- **Barra de herramienta:** Es la parte en la cual nos permite seleccionar parte de la imagen (píxeles), para un análisis enfocado, tanto en la imagen, como en el espectro. Cuenta con herramienta de selección tipo lazo, seleccionador de cuadrantes, seleccionador tipo *mira*, el cual recoge píxel por píxel.
- **Panel de imagen:** Donde se visualiza la imagen de una longitud de onda por cada píxel. Este panel no visualiza el hipercono, pero sí las diferentes capas de él, al modificar los falsos colores.
- **Panel de trazos:** Este panel nos permite ver el espectro de 900 a 1700 nm de cada píxel señalado por la herramienta *mira*, o por las áreas seleccionadas (más de un píxel). En ella se puede ver el brillo que refleja en cada píxel de imagen para los 168 canales.

- Panel de herramienta: Es el panel que nos permite, cambiar en un falso color a nuestra imagen. Cambiar los contrastes y poder ver las diferentes capas con un falso color.
- Panel de datos: En él se puede observar la información de cada píxel y la ubicación en el plano con su respectiva longitud de onda.
- Panel de árbol de recursos : En ese panel se puede observar las diferentes opciones para diversas imágenes hiperespectrales, tanto en imagen y espectro, en la cual se encontrarán opciones de poder ver la media de un conjunto de píxel, por ende sus espectros medio.
- Panel de información: En él se observa toda la información de la imagen y del espectro. En él se puede ver todos los espectros vs brillo de un espectro.

Inicio de SpectrononPro

Inicie SpectrononPro haciendo doble click en el ícono de SpectrononPro, o desplácese hasta él desde su menú de Inicio. Desde el menú principal, seleccione Archivo → Preferencias ... Esto revelará el menú de Preferencias. Seleccione la pestaña de su espectrómetro de imagen Pika.



Figura 3.66: Icono de Spectronon Pro.

Calibración del sistema hiperespectral

Se utiliza para reducir el efecto de la distribución no homogénea de la luz en la cámara. Esta corrección se usa imágenes de referencia oscura (RD) y blanco (RW). El RW se adquiere de un patrón de teflón, con valor de reflectancia $\sim 99.9\%$ y la RD se adquiere al cubrir la lente con una tapa oscura, reflectancia $\sim 0.0\%$. La corrección de las imágenes se realiza mediante la siguiente ecuación.

$$R_c = (R - R_D) / (R_W - R_D) \quad (16)$$

Donde:

- R_c = HSI corregida en reflectancia relativa
- R = HSI en bruto
- R_D = HSI en negro

- R_W = HSI de referencia blanca

para ello en el sistema se debe calibrar con los siguientes pasos:

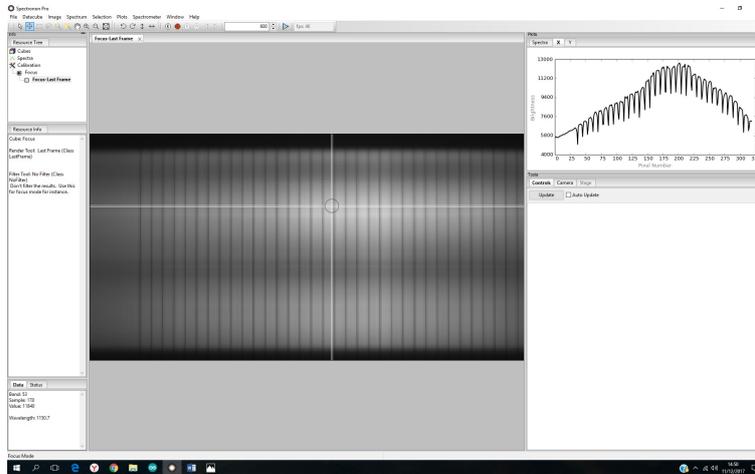


Figura 3.67: Calibración del foco.

En este paso se debe calibrar el foco, para lo cual debemos poner en la mesa de trabajo la hoja de calibración, y obtener varias líneas, que deben estar enfocadas y el espectro debe ser continuo, así como se observa en la figura 3.67.

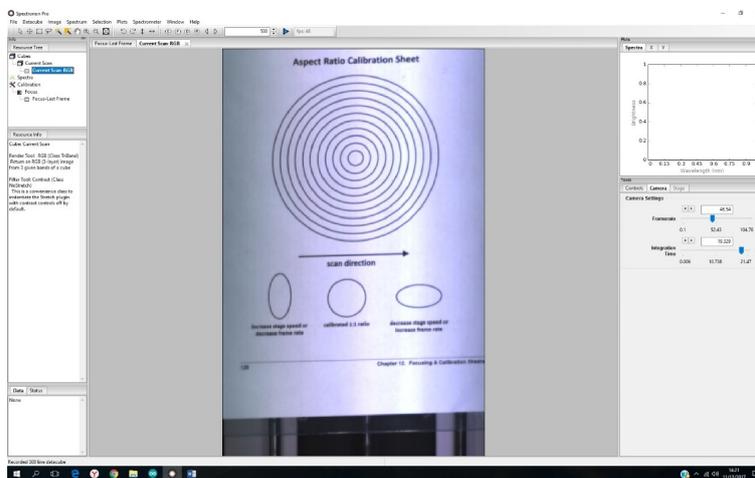


Figura 3.68: Calibración de la velocidad de la mesa.

Para la siguiente calibración se regula el movimiento de la mesa (tiempo de integración) y los fotogramas que realiza la cámara. Para ello se pone una hoja de calibración, como se observa en la imagen 3.68, para ello se debe visualizar un círculo en la imagen.



Figura 3.69: Calibración de referencia oscura y blanca.

Se prosigue calibrando un fondo de radiación oscura (referencia oscura). Para ello se cubre el lente y se hace click en *D* en la barra de herramienta. Se prosigue calibrando una referencia blanca, para ello se coloca un teflon patrón y se hace click en *R* en la barra de herramienta.

Adquisición de imágenes

Se procede a dar click en el botón play que aparece en la barra de herramienta, simultáneamente se hace deslizar la mesa de trabajo, ya antes calibrada.



Figura 3.70: Hojas y flores.

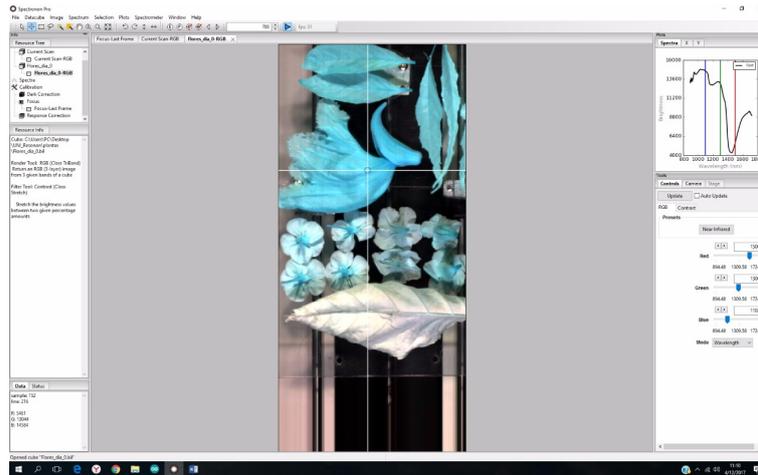


Figura 3.71: Imagen tomada con la cámara Resono PIKA-NIR.

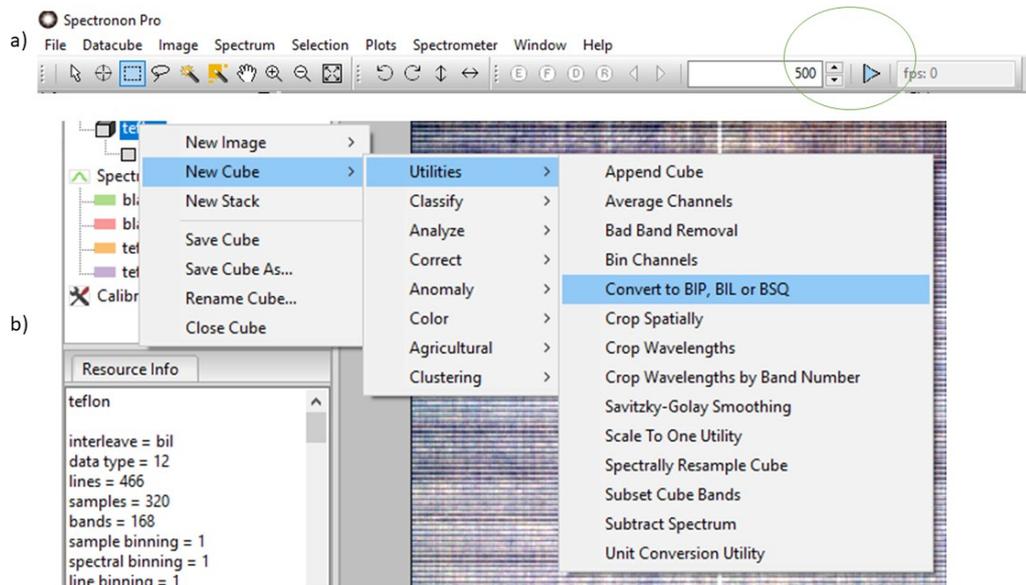


Figura 3.72: a) barra de herramienta, inicio de cámara b) utilidad del software para pasar de .BIL a .BSQ.

Anexo B Código [R]

Ricardo Flores

27/6/2019

Para instalar alguna librería en [R] y abrirlas, se escribe los siguientes comandos

```
install.packages("ggplot2")  
library(ggplot2)
```

Las librerías que se utilizan son las siguientes:

```
library( hyperSpec) #Permita abrir Los datos con formato .bil  
library( signal)   #Permita realizar Los filtros savitzky y derivadas  
library( pls )     #  
library( prospectr) #SNV y detrend  
library( fields)   #Permite ver imagenes (image.plot)  
library( fdth)     #Realiza La distribución de frecuencia, para hacer um  
bralización  
library(readr)     #ermite Leer datos txt o csv  
library( ggplot2)  
library( rgdal)  
library( raster)  
library( caret)  
library( dplyr)  
library( caTools)  
library( lattice)  
library( grid)  
library( gplots)
```

(1) Abriendo Datos

Los archivos con extensión .bil que se obtienen del software Spectronon-PRO se guardan en la carpeta que por defecto crea [R] y que se puede ver con el comando **getwd()**. Se crearon tres vectores que contiene las imágenes hiperespectrales con los nombres: derivados, suelos crudo y suelos diésel. De esta manera se pueden procesar grandes cantidades de datos. Las imágenes hiperespectrales utilizadas se pueden ver en el siguiente script.

```
# derivados -----  
--  
muestra_1='crudo.bil'  
muestra_2='diesel.bil'  
muestra_3='gasolina.bil'  
# suelos -----  
--  
muestra_4='\\limosa.bil'
```

```

muestra_5='\\arcilla.bil'
muestra_6='\\arenosa.bil'
derivados<-c(muestra_1,muestra_2,muestra_3,muestra_4,muestra_5,muestra_6)

# suelos_crudo -----
--
muestra_8='lim_cru_0.bil'
muestra_9='lim_cru_3.bil'
muestra_10='lim_cru_6.bil'
muestra_11='lim_cru_9.bil'
muestra_12='lim_cru_12.bil'
muestra_13='lim_cru_15.bil'
muestra_14='lim_cru_18.bil'
muestra_15='lim_cru_21.bil'
muestra_16='lim_cru_24.bil'
muestra_17='lim_cru27.bil'
muestra_18='lim_cru_30.bil'
muestra_arc='Arc_lim_00_200.bil'
muestra_19='Arc_lim_03_200.bil'
muestra_20='Arc_lim_06_200.bil'
muestra_21='Arc_lim_09_200.bil'
muestra_22='Arc_lim_12_200.bil'
muestra_23='Arc_lim_15_200.bil'
muestra_24='Arc_lim_18_200.bil'
muestra_25='Arc_lim_21_200.bil'
muestra_26='Arc_lim_24_200.bil'
muestra_27='Arc_lim_27_200.bil'
muestra_28='Arc_lim_30_200.bil'
muestra_are='are_cru_0.bil'
muestra_29='are_cru_2.bil'
muestra_30='are_cru_4.bil'
muestra_31='are_cru_6.bil'
muestra_32='are_cru_8.bil'
muestra_33='are_cru_10.bil'
muestra_34='are_cru_12.bil'
muestra_35='are_cru_14.bil'
muestra_36='are_cru_16.bil'
muestra_37='are_cru_18.bil'
muestra_38='are_cru_20.bil'
suelos_crudo<-c(muestra_8,muestra_9,muestra_10,muestra_11,muestra_12,mues
tra_13,muestra_14,
                muestra_15,muestra_16,muestra_17,muestra_18,muestra_arc,m
uestra_19,muestra_20,muestra_21,muestra_22,
                muestra_23,muestra_24,muestra_25,muestra_26,muestra_27,mu
estra_28,muestra_are,muestra_29,muestra_30,
                muestra_31,muestra_32,muestra_33,muestra_34,muestra_34,mu
estra_35,muestra_36,
                muestra_37,muestra_38)

# suelos_diesel -----

```

```

--
muestra_lim='lim_cru_0.bil'
muestra_39='lim_dis_3.bil'
muestra_40='lim_dis_6.bil'
muestra_41='lim_dis_9.bil'
muestra_42='lim_dis_12.bil'
muestra_43='lim_dis_15.bil'
muestra_44='lim_dis_18.bil'
muestra_45='lim_dis_21.bil'
muestra_46='lim_dis_24.bil'
muestra_47='lim_dis_27.bil'
muestra_48='lim_dis_30.bil'
muestra_arc='Arc_lim_00_200.bil'
muestra_49='arc_dis_3.bil'
muestra_50='arc_dis_6.bil'
muestra_51='arc_dis_9.bil'
muestra_52='arc_dis_12.bil'
muestra_53='arc_dis_15.bil'
muestra_54='arc_dis_18.bil'
muestra_55='arc_dis_21.bil'
muestra_56='arc_dis_24.bil'
muestra_57='arc_dis_27.bil'
muestra_58='arc_dis_30.bil'
muestra_are='are_cru_0.bil'
muestra_59='are_die_2.bil'
muestra_60='are_die_4.bil'
muestra_61='are_di_6.bil'
muestra_62='are_di_8.bil'
muestra_63='are_die_10.bil'
muestra_64='are_die_12.bil'
muestra_65='are_die_14.bil'
muestra_66='are_die_16.bil'
muestra_67='are_die_18.bil'
muestra_68='are_die_20.bil'
suelos_diesel<-c(muestra_lim,muestra_39,muestra_40,muestra_41,muestra_42,
muestra_43,
muestra_44,muestra_45,muestra_46,muestra_47,muestra_48,m
uestra_arc,
muestra_49,
muestra_50,muestra_51,muestra_52,muestra_53,muestra_54,m
uestra_55,
muestra_56,muestra_57,muestra_58,muestra_are,muestra_59,
muestra_60,muestra_61,
muestra_62,muestra_63,muestra_64,muestra_65,muestra_66,m
uestra_67,
muestra_68)

```

Para abrir cada uno de las imágenes, se hace una rutina que utiliza los vectores antes creados; pero antes de todo se crea algunos array con valores NA que se utiliza para escribir los valores que se van a determinar.

```

#Llamamos una librería que nos permitirá abrir un archivo txt que contiene los valores de longitud de onda.
library(readr)
long <- read_csv("C:/Users/USUARIO/Desktop/long.txt")
long<-as.data.frame(long)
longitud<-long[15:150,1]
View(longitud)
#Se crea los arrays que permitirán ser utilizados para llenarlos con los valores que encontremos.
total_matrix<-array(NA,c(40000,136,33))
total1_matrix<-array(NA,c(40000,136,33))
total2_matrix<-array(NA,c(40000,136,33))
#View(Longitud)
pixe_too<-array(NA,c(33))
cubo_sf<-array(NA,c(1000,136,33))
total<-array(NA,c(136,33))
total1<-array(NA,c(136,33))
total2<-array(NA,c(136,33))
total3<-array(NA,c(136,33))
intex_to<-array(NA,c(33,4))
#proc.time nos permite calcular el tiempo que demora nuestro script
t <- proc.time()
#en esta parte empieza la iteración para un vector de 33 variables: ejm suelo_crudo
# de la misma manera ponemos suelo_diesel y derivados.
for (k in 1:33) {
  k
  #con este código se abren los archivos .bil
  arcilla_cubo=hyperSpec::read.ENVI(suelos_crudo[k])
  #se pasa de un cubo a una matriz
  arcilla_matriz<-as.matrix(arcilla_cubo,byrrow=T)
  #solo se escoge 136 bandas
  arcilla_matriz_136band<-arcilla_matriz[,c(15:150)]
  bandx<-sqrt(dim(arcilla_matriz_136band)[1])

```

(2) Pre-Procesamiento

Dentro del script y la iteración se encuentra el preprocesamiento, para ello se creó una función llamada Preprocesamiento, la cual se mostrará finalizando la iteración.

```

# preprocesamiento -----
-----
#función para el preprocesamiento
source("preprocesamiento.R")
resultad<-preprocesamiento(arcilla_matriz_136band,136)
names(resultad)

```

(3) Reducción de dimensiones

En la reducción de dimensiones se utiliza el análisis de componentes principales (PCA), y umbralización (threshold). De la misma manera que se creó una función para el Preprocesamiento, también se crea una función para la reducción de dimensiones.

```
# PCA -----
-----
#función para la reducción de dimensiones
source("pca_mio.R")
pco<-pca_mio(resultad$SG,bandx,bandx)
dev.off()
names(pco)
pco$pc1_a
pc_1_N<-pco$pc1_a
#CRUDO
N1<-c(-4,-4,-3,-3,-3,-3,-4,-3,-3,-4,-4,-3,-3,-2,-3,-3,-3,-4,-5,-5,-5,-5,
-5,-4,-4,-4,-4,-3,-4,-4,-4,-4,-3)

N2<-c(-1,-1,-1,-1,-1,-1,-1,0,0,0,1,0,0,0,-1,-1,-1,-1,-1,-1,-1,-3,-2,-2,
-2,-2,-2,-1,-2,-2,-2,-2,-1)
#
# # Diesel
# N1<-c(-4,-4,-3,-3,-3,-3,-3,-3,-4,-4,-4,-3,-3,-3,-4,-3,-4,-5,-4,-4,-4,
-4,-5,-5,-5,-5,-5,-5,-5,-5,-5,-2,-2)
#
# N2<-c(0,0,-1,0,0,0,0.5,0,0,0,0.5,0,0,0,0,0,0,-1,-1,-2,-1,-1,-2,-2,-2,
-2,-2,-2.5,-2.5,-2.5,-2,0,0)
```

continuando la creación de mascara y salidas de los datos preprocesados: Nor/SV-g, Nor/SV-g/SNV y Nor/SV-g/SNV-D

```
pixx<-200
for (i in 1:pixx) {
  for (j in 1:pixx) {
    if (pc_1_N[i,j]<N1[k]) {pc_1_N[i,j]=N2[k]+10}
    if (pc_1_N[i,j]>N2[k]){pc_1_N[i,j]<-NA}
  }
}

#hist(pc_1_N,breaks = 150)
# image.plot(t(pc_1_N))
pc_m1<-as.vector(pc_1_N)
#boxplot(pc_m1,horizontal = T)
#summary(pc_m1)
RI<-abs(summary(pc_m1)[5]-summary(pc_m1)[2])
inf<- summary(pc_m1)[2]-1.5*RI #summary(pc_m1)[2]
sup<-summary(pc_m1)[5]+1.5*RI#summary(pc_m1)[5]
source("reduccion_dime.R")
```

```

S_G<-reduccion_dim(pco$pc1_a,resultad$SG,inf,sup,136,bandx,bandx)
dev.off()
source("reduccion_dime.R")
S_N_V<-reduccion_dim(pco$pc1_a,resultad$SNV11,inf,sup,136,bandx,bandx)
dev.off()
source("reduccion_dime.R")
S_N_D<-reduccion_dim(pco$pc1_a,resultad$SND11,inf,sup,136,bandx,bandx)
names(S_N_D)
dev.off()
#matr_30: promedio
total[,k]<-S_G$matr_30
total1[,k]<-S_N_V$matr_30
total2[,k]<-S_N_D$matr_30
#matr_3: matrix de todo Los pixeles
total_matrix[, ,k]<-S_G$matr_3
total1_matrix[, ,k]<-S_N_V$matr_3
total2_matrix[, ,k]<-S_N_D$matr_3

}
proc.time() - t

```

(3.1) función: preprocesamiento

```

# solo preprocesamiento
getwd()
preprocesamiento<-function(A,Num_band){
#rw=nrow(A);cl=ncol(A);mx=max(A);mi=min(A)
#A_nor1=array(0,c(rw,cl))
#for (i in 1:rw) {
# for (j in 1:cl){
#   A_nor1[i,j]=A[i,j]/mx
# }
#}
#A_nor1 ;dim(A_nor1)
#B=A_nor1
#A se elimino para que no se normalicen, y defrente se hace el golay
B=A
library(signal)#para savatzky
rw=nrow(B);cl=ncol(B)
sg_res1=array(0,c(rw,cl))
for (i in 1:rw) {
  sg_res1[i,]=signal::filter(filt = sgolay(5,47,0), x =B[i,])#sgolay(7,25
,0) #si=sgolay(5,47,0)
}
Xt<-t(sg_res1)
Xt_snv<-scale(Xt,center=TRUE,scale=TRUE)
res1d<-t(Xt_snv)
library(pls)
library(prospectr)
x_snv<-standardNormalVariate(res1d)
wav<-c(1:Num_band))

```

```

x_dt<-detrend(sg_res1,wav)
SG<-sg_res1;SG
SNV11<-x_snv;SNV11
SND11<-x_dt;SND11
return(list(SG=SG,SNV11=SNV11,SND11=SND11))
print("output: SG,SNV11,SND11")
}

```

(3.2) función: Reducción de dimensiones

```

reduccion_dim<-function(x4,matr_3,N1,N2,Num_band,pixx1,pixx2){
  pixx<-pixx1<-pixx2
  for (i in 1:pixx) {
    for (j in 1:pixx) {
      if (x4[i,j]<N1) {x4[i,j]=N2*10}
      if (x4[i,j]>N2){x4[i,j]<-NA}
    }
  }
  imagen_x1<-x4
  pixel=pixx*pixx
  res<-imagen_x1
  Nuevo_res<-array(res,dim=c(pixel))
  Nuevo_res[is.na(Nuevo_res)]<-2000
  for (i in 1:pixel) {
    if (Nuevo_res[i]<999) {
      Nuevo_res[i]<-NA
    }else{
      Nuevo_res[i]<-Nuevo_res[i]
    }
  }
  Nuevo_res<-data.frame(c(1:pixel),Nuevo_res) #con 16184 "NA"
  S_Na_res<-na.omit(Nuevo_res)
  dimension<-dim(S_Na_res[,2]);dimension
  pixeles<-S_Na_res[,1]
  #se ve la muestra cubo
  # matr_2<-sg_res1
  # for (i in pixeles) {
  #   matr_2[i,]<-NA #rnorm(1,10,1)
  #
  # }
  for (i in pixeles) {
    matr_3[i,]<-NA
  }
  datos1<-array(matr_3,c(pixx,pixx,136))
  matr_30<-colMeans(na.omit(matr_3))
  dev.new()
  par(mfrow=c(1,3))
  image.plot(imagen_x1)
  image.plot(datos1[,4],main="bandas")
  matplot(c(1:136),(matr_30),"l")
}

```

```

return(list(matr_3=matr_3,matr_30=matr_30))
}

```

(4) Salida de datos

Gráficas, espectros, elección de banda de interés, segunda derivada y regresión lineal

```

lim<-c(1:11)
suelos<-c("limoso","arcilloso","arenoso");k=1
phc<-c("crudo","diésel","gasolina");j=1
nomb<-c("long","0%","3%","6%",
        "9%","12%","15%",
        "18%","21%","24%",
        "27%","30%");i=12
nomb<-c("long","0%","2%","4%",
        "6%","8%","10%",
        "12%","14%","16%",
        "18%","20%");i=12
library(reshape2)
dat<-data.frame(longitud,total[,lim])
#View(dat)
colnames(dat)<-nomb[1:i]
dc<-melt(dat, id="long")#dat[,c(1,5,6,7)]
dev.new()
library(ggplot2)
ggplot(dc, aes(x=long, y=value, color=variable)) +
  geom_line(aes(linetype=variable), size=1)+
  labs(
    title =paste(suelos[k],"-", phc[j],sep = "" ) ,
    x="longitud de onda (nm)",
    y=" Reflectancia ",
    caption = "suavizado savitzky-Golay",size=2
  )+
  theme_classic()+
  scale_x_continuous(breaks = seq(900,1700,100))

#SNV
dat<-data.frame(longitud,total1[,lim])
colnames(dat)<-nomb[1:i]
dc<-melt(dat, id="long")#dat[,c(1,5,6,7)]
dev.new()
library(ggplot2)
ggplot(dc, aes(x=long, y=value, color=variable)) +
  geom_line(aes(linetype=variable), size=1)+
  labs(
    title =paste(suelos[k],"-", phc[j],sep = "" ) ,
    x="longitud de onda (nm)",
    y="Reflectancia SNV",
    caption = "suavizado savitzky-Golay/SNV",size=2
  )+

```

```

theme_classic()+
scale_x_continuous(breaks = seq(900,1700,100))

#SNV -D
dat<-data.frame(longitud,total2[,lim])
colnames(dat)<-nomb[1:i]
dc<-melt(dat[c(27:85),], id="long")#dat[,c(1,5,6,7)]
dev.new()
library(ggplot2)
ggplot(dc, aes(x=long, y=value, color=variable)) +
  geom_line(aes(linetype=variable), size=1)+
  labs(
    title =paste(suelos[k],"-", phc[j],sep = ""),
    x="longitud de onda (nm)",
    y=" Reflectancia SNV-D",
    caption = "suavizado savitzky-Golay/SNV-D",size=2
  )+
  theme_classic()+
  scale_x_continuous(breaks = seq(900,1700,50))

#derivada
a3<-total2
dat<-data.frame(longitud,a3[,lim])
SV<-dat[1:136,]
for (i in 2:12) {
  SV[,i]<-signal::filter(filt = sgolay(5,25,2), x =dat[c(1:136),][,i])
}
dat<-SV
colnames(dat)<-nomb[1:i]
dc<-melt(dat[c(36:65),1:12], id="long")#dat[,c(1,5,6,7)]
dev.new()
library(ggplot2)
ggplot(dc, aes(x=long, y=value, color=variable)) +
  geom_line(aes(linetype=variable), size=1)+
  labs(
    title =paste(suelos[k],"-", phc[j],sep = ""),
    x="longitud de onda (nm)",
    y=" segunda derivada Reflectancia SNV-D",
    caption = "suavizado savitzky-Golay/SNV-D/2d",size=2
  )+
  theme_classic()+
  scale_x_continuous(breaks = seq(900,1700,20))

# La derivada
library(dplyr)
library(highcharter)
hchart(dc, "line", hcaes(x =long,y=value,group=variable)) %>%
  hc_add_theme(hc_theme_google())

```

```

#derivada con tres puntos
a2<-array(NA,c(136,33))
for (i in 1:11) {
  for (k in 1:3) {
    num<-dim(na.omit(total2_matrix[, ,i])[,])[1]
    z=2.58;e=0.05;N=num;sigma=0.5
    n2<-(z^2)*(sigma^2)*(N)/((e^2)*(N-1)+(z^2)*(sigma^2)      )
    n22<-trunc(n2) #entero
    vals<-sample(num,size = n22)
    a1<-na.omit(total2_matrix[, ,i])[vals,]
    a2[,c(3*(i-1)+k)]<-colMeans(a1)
  }
}
a2
a3<-as.data.frame(a2)

#a3<-total2
lim=33
dat<-data.frame(longitud,a3[,])

SV<-dat[1:136,]
for (i in 2:34) {
  SV[,i]<-signal::filter(filt = sgolay(5,25,2), x =dat[c(1:136),][,i])
}
dat<-SV

View(longitud)
df<-as.numeric(dat[50,c(2:31)])
#df<-as.numeric(abs(dat[40,c(2:8)]-dat[56,c(2:8)]))
conce<-c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,
        6,6,6,7,7,7,8,8,8,9,9,9,
        10,10,10)
#conce<-c(0,3,6,9,12,15,18,21,24,27,30)
#df<-med[c(50),c(1:11)]/med[c(50),11]
cv<-lm(conce~df)
sunti<-paste("wt%=",round(cv$coefficients[1],3),"+",
             round(cv$coefficients[2],3), "*R' ' ", " r2=",
             round(as.numeric(summary(cv)[8]),3),sep = "" )

dev.new()
plot(df,conce,
     main=paste(suelos[k],"-", phc[j],sep = ""),
     xlab=expression(dR^2/d^2~lambda),
     ylab="concentración %",bty='n',pch=16,col="blue"
)
legend("topleft",legend=sunti,pch=19,col=c(1,2,3),bty='n')
abline(cv,col="red")

```

```

#con sd
i=1
b1<-colMeans(total2_matrix[,i],na.rm = T)
b2<-apply(total2_matrix[,i],2,sd,na.rm=T)
b3<-na.omit(total2_matrix[,1])

dim((b3))[1]
SV<-t(b3)
for (i in 1:dim((b3))[1]) {
  SV[,i]<-signal::filter(filt = sgolay(5,25,2), x =t(b3)[,i])
}
dat<-SV

```

(5) Clasificación de las imágenes HSI

```

h_S_G<-total_matrix
#limosa
rs_0<-rbind(h_S_G[,1],h_S_G[,2],h_S_G[,3],h_S_G[,4]);rs_1<-rbind(h_S_G[,5],h_S_G[,6],h_S_G[,7],h_S_G[,8]);blanco<-array(NA,c(40000,136));rs_2<-rbind(h_S_G[,9],h_S_G[,10],h_S_G[,11],blanco)

#arcillosa
rs_0<-rbind(h_S_G[,12],h_S_G[,13],h_S_G[,14],h_S_G[,15]);rs_1<-rbind(h_S_G[,16],h_S_G[,17],h_S_G[,18],h_S_G[,19]);blanco<-array(NA,c(40000,136));rs_2<-rbind(h_S_G[,20],h_S_G[,21],h_S_G[,22],blanco)

#arenosa
rs_0<-rbind(h_S_G[,23],h_S_G[,24],h_S_G[,25],h_S_G[,26]);rs_1<-rbind(h_S_G[,27],h_S_G[,28],h_S_G[,29],h_S_G[,30]);blanco<-array(NA,c(40000,136));rs_2<-rbind(h_S_G[,31],h_S_G[,32],h_S_G[,33],blanco)
#res1_3Dm<-array(rs_27_30,dim=c(200,200*4,136))

nr<-800 ; nc<-200; nw=136
img1=array(0,c(nr,nc,nw))
img2=array(0,c(nr,nc,nw))
img3=array(0,c(nr,nc,nw))
for ( i in 1:nr) {
  for ( j in 1:nc) {
    n = nc*(i-1)+j #( j - 1) * nc + i
    img1 [i , j , ]=rs_0[n, ]
    img2 [i , j , ]=rs_1[n, ]
    img3 [i , j , ]=rs_2[n, ]
  }
}

img1_a<-array(img1,c(160000,136))
img2_a<-array(img2,c(160000,136))
img3_a<-array(img3,c(160000,136))
dac<-rbind(img1_a,img2_a,img3_a)

```

```

resc<-array(dac,dim = c(800,600,136))

dac[is.na(dac)]<-10
osos<-sample(c(140:160),136,replace = T)/100
for (j in 1:136) {
  for (i in 1:480000) {
    if (dac[i,j]>5) {dac[i,j]<-rnorm(1,osos[j],0.2) }
  }
}

res1_PC<-princomp(dac,scores=TRUE)
prop_var<-res1_PC$sdev^2/sum(res1_PC$sdev^2)
prop_var[1]

pc1<-res1_PC$scores[,1]

pc_c_m<-array(pc1,c(800,600))

for (i in 1:800) {
  for (j in 1:600) {
    if (pc_c_m[i,j]>2) {pc_c_m[i,j]<-NA}
  }
}

name1<-paste("PC1-Scores ",round(prop_var[1],4)*100,"%",sep = "")
dev.new()
image.plot(pc_c_m,axes=F,main=name1)

# densidades y tabla -----
--

cubito<-array(NA,c(200,200,11))
xdv<-array(NA,c(11,10))

for (i in 1:4) {
  for (j in 1:3) {
    xx1<-1+200*(i-1);xx1
    xx2<-200*i;xx2
    yy1<-1+200*(j-1)
    yy2<-200*j

    xdv[j+3*(i-1),]<-c(i,j,summary(as.vector(pc_c_m[xx1:xx2,yy1:yy2])),sd
(pc_c_m[xx1:xx2,yy1:yy2],na.rm = T))
    cubito[,j+3*(i-1)]<-pc_c_m[xx1:xx2,yy1:yy2]
    #image.plot(t(cubito[,j+3*(i-1)]),axes=F)
  }
}

```

```

dev.new()
hist(na.omit(as.vector(cubito[,1])),breaks = "Sturges",freq = F,
      ylim=c(0,2),xlim =c(-11,-4) ,main = "xcv",
      xlab="Scores",ylab = "densidad")
for (i in c(1,4,7,10,2,5,8,11,3,6,9)) {

  eve<-na.omit(as.vector(cubito[,i]))
  curve(dnorm(x,mean = mean(eve),sd =sd(eve) ),col=i,add = T,lwd=2)

}

# hacer tabla -----
--
listabien<-data.frame((as.vector(cubito[,1])),(as.vector(cubito[,4])),
                      (as.vector(cubito[,7])),(as.vector(cubito[,10])),
                      (as.vector(cubito[,2])),(as.vector(cubito[,5])),
                      (as.vector(cubito[,8])),(as.vector(cubito[,11])),
                      (as.vector(cubito[,3])),(as.vector(cubito[,6])),
                      (as.vector(cubito[,9])))

distancia<-c(abs(mean(listabien[,1],na.rm = T)-mean(listabien[,2],na.rm =
T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,3],na.rm =
T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,4],na.rm =
T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,5],na.rm =
T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,6],na.rm =
T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,7],na.rm =
T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,8],na.rm =
T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,9],na.rm =
T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,10],na.rm
= T)),
             abs(mean(listabien[,1],na.rm = T)-mean(listabien[,11],na.rm
= T)))

concentracion<-c(3,6,9,12,15,18,21,24,27,30)

prob_suelo<-array(NA,c(10))
prob_mezcla<-array(NA,c(10))
limite<-array(NA,c(10))
media_suelo<-array(NA,c(10))
media_mezcla<-array(NA,c(10))

```

```

sd_suelo<-array(NA,c(10))
sd_mezcla<-array(NA,c(10))
pixeles1<-array(NA,c(10))
pixeles2<-array(NA,c(10))
for (i in 1:10) {

  m1<-mean(listabien[,1],na.rm = T)
  m2<-mean(listabien[,i+1],na.rm = T)
  sd1<-sd(listabien[,1],na.rm = T)
  sd2<-sd(listabien[,i+1],na.rm = T)
  pix1<-40000-summary(listabien[,1])[7]
  pix2<-40000-summary(listabien[,i+1])[7]
  limite[i]<-(m2+m1)/2
  prob_suelo[i]<-1-pnorm((m2+m1)/2,mean = m1,sd =sd1)
  prob_mezcla[i]<-pnorm((m2+m1)/2,mean = m2,sd =sd2)
  media_suelo[i]<-m1
  media_mezcla[i]<-m2
  sd_suelo[i]<-sd1
  sd_mezcla[i]<-sd2
  pixeles1[i]<-pix1
  pixeles2[i]<-pix2
}

caracteris<-data.frame(concentracion,media_suelo,sd_suelo,media_mezcla,sd
_mezcla,
                      distancia,limite,prob_suelo,prob_mezcla,pixeles1,p
ixeles2)
View(caracteris)

names(caracteris)

caracteris$sd_distancia<-sqrt(caracteris$sd_suelo^2 + caracteris$sd_mezcl
a^2)

x<-caracteris$concentracion
y<-caracteris$distancia
sd<-caracteris$sd_distancia
dev.new()
qplot(x,y)+
  geom_errorbar(aes(x=x, ymin=y-sd, ymax=y+sd),width=0.25)

```