



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS TABASCO

PROGRAMA EN PRODUCCIÓN AGROALIMENTARIA EN EL TRÓPICO

Estudio de la Contaminación con Petróleo de los Suelos Agrícolas de la Cuenca Baja del río Tonalá Mediante Espectroscopia del Infrarrojo Cercano

LORENA DEL CARMEN HERNÁNDEZ NATAREN

TESIS

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE: MAESTRO EN CIENCIAS

H. CÁRDENAS, TABASCO

Noviembre 2009

La presente tesis titulada: **Estudio de la Contaminación con Petróleo de los Suelos Agrícolas de la Cuenca Baja del río Tonalá Mediante Espectroscopia del Infrarrojo Cercano**, realizada por la alumna Lorena del Carmen Hernández Nataren, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS

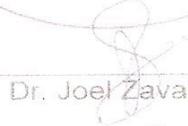
PRODUCCION AGROALIMENTARIA EN EL TROPICO

CONSEJO PARTICULAR

CONSEJERO:


Dr. Armando Guerrero Peña

ASESOR:


Dr. Joel Zavala Cruz

ASESOR:


Dr. Eugenio Carrillo Ávila

ASESOR:

Dr. María del Carmen Rivera Cruz

Cárdenas, Tabasco, México, 23 de Noviembre de 2009

Resumen

El desarrollo de la industria petrolera en México ha generado contaminación en el entorno, ocasionada por derrames que afectan componentes del ecosistema donde se localiza su infraestructura. En la actualidad se desarrollan métodos modernos para estudiar la contaminación de suelos generada por el derrame de hidrocarburos derivados del petróleo.

Los métodos instrumentales modernos (espectroscopia de infrarrojo cercano, cromatografía de gases, espectrometría de masas, espectroscopia UV-Vis, etc.), presentan gran interés principalmente por su elevada velocidad de análisis, por la reducción de los errores analíticos y en el consumo de reactivos, entre otros.

El objetivo del trabajo es generar modelos de clasificación y estimación de la contaminación en suelos agrícolas de la Cuenca Baja del río Tonalá con hidrocarburos derivados del petróleo mediante la tecnología NIRS. Para ello se utilizaron cinco grupos de suelos (76 muestras) con diferentes características físicas y químicas. Los resultados obtenidos se trataron quimiométricamente, aplicando técnicas para visualizar la distribución de los datos y la eliminación de la región visible del espectro NIRS. Posteriormente se generó el modelo de clasificación mediante PCA y el modelado de clases con SIMCA. Los modelos se validaron interna y externamente con valores obtenidos con un método de referencia (Soxhlet) presentando valores de clasificación y estimación que permiten ver la tendencia para estimar la contaminación de suelos. Por lo tanto, la tecnología de infrarrojo cercano genera información suficiente para estudiar la contaminación de suelos por hidrocarburos derivados del petróleo con resultados similares a los obtenidos en los métodos convencionales.

Abstract

In Mexico there are large areas contaminated with hydrocarbons of petroleum, due to the own activities of the oil industry, as well as to accidental spills exist mainly during the extraction and/or transports. The hydrocarbons derived from petroleum, are accumulated in ecosystems marine and grounds, being responsible for their degradation, reason for which modern methods are developed to evaluate the soil pollution generated by the oil spills.

Nowadays, the development of new technologies and modern instruments such as Near Infrared Spectroscopy (NIRS), mass spectrometry (MS), Gas Chromatography (GC), among others, are considered a significant advance for this type of study mainly due to the speed of analysis, analytical error reduction and decreased consumption of reagents.

The objective was to generate classification models and estimation of pollution in agricultural soils by hydrocarbons derived from the petroleum from the Tonalá River Lower using NIRS technology on five units of soils (76 samples) with different physical and chemical characteristics. The results were treated chemometric, with techniques to visualize the data distribution and elimination of the visible spectrum NIRS.

Subsequently, we generated a classification model using PCA and modeling classes with SIMCA. The models were validated internally and externally with values obtained with a reference method (Soxhlet) presenting values of classification and estimation that can see the trend to estimate the contamination of soils. Therefore, the technology of near infrared generated sufficient information to study soil pollution by hydrocarbons derived from petroleum with similar results to those obtained by conventional methods.

Dedicatoria

A Dios porque en su infinita misericordia y paciencia, me ha prestado la vida todo este tiempo, y gracias a ello puedo lograr mis metas.

A mis padres y hermanos quienes están conmigo siempre, y por ser el motor que mueve mi vida.

A mi cuñada y sobrinos, los cuales siempre me motivan a hacer lo que más me gusta.

A todos y cada uno de ustedes los tengo siempre presente, pues son quienes me dan la fuerza, y las ganas de seguir adelante sin importar cuán difícil sea el camino.

Agradecimiento

Al Dr. Armando Guerrero Peña, por su amistad, apoyo y el tiempo dedicado a este proyecto.

A los Doctores Joel Zavala Cruz, Eugenio Carrillo Ávila y la doctora María del C. Rivera Cruz, asesores e integrantes de mi Consejo Particular, por el tiempo y el apoyo brindado.

A la línea prioritaria de investigación 1: Manejo Sustentable de Recursos Naturales (MSNR) por su apoyo a través del proyecto Estudio integral de la Cuenca Hidrológica del río Tonalá y Lagunas del Carmen y Machona.

A Edith y Aarón por su infinito apoyo, en todo lo que he realizado.

A todos y cada uno de las personas que han estado conmigo cuando los he necesitado.

Al personal del Laboratorio de Suelos (LASPA) del Colegio de Postgraduados Campus Tabasco, (Martha, Bernardo, Esteban, Emilgar, Gina y María) por su amistad y apoyo.

A los millones de mexicanos (as) que pagan impuestos, quienes, a través del Consejo Nacional de Ciencia y Tecnología (CONACYT) y el Colegio de Postgraduados, me han permitido hacer lo que más me gusta.

A todos y cada uno de ustedes MIL GRACIAS.

Índice

Índice de cuadros.....	xi
Índice de figuras	xiii
1. INTRODUCCION	1
2. OBJETIVOS	3
2.1. Objetivo General	3
2.1.1. Objetivos Específicos.....	3
3. HIPÓTESIS	4
3.1. Hipótesis General.....	4
3.1.1. Hipótesis específicas	4
4. REVISIÓN BIBLIOGRÁFICA.....	5
4.1. La Contaminación con Petróleo Crudo.....	5
4.2. Persistencia y Fraccionamiento de los Hidrocarburos Derivados del Petróleo en el Suelo	7
4.3. Normativas Ambientales.....	7
4.4. Composición Química y Propiedades Físicas del petróleo Crudo	8
4.4.1. Hidrocarburos derivados del petróleo	9
4.4.1.1. Compuestos saturados e insaturados.....	10
4.5. Técnicas para la Cuantificación de Hidrocarburos del petróleo en Suelos	10
4.5.1. Métodos convencionales	10
4.5.2. Métodos modernos	12
4.5.2.1. Olfatometría electrónica	13
4.5.2.2. Equipos basados en espectrometría de masas	13
4.5.2.3. Espectroscopía en el infrarrojo cercano (NIRS).....	13
4.5.3. Aplicaciones de la tecnología NIRS.	14

4.6. Teoría de la Absorción en el Infrarrojo	15
4.7. Interpretación de Espectros.....	19
4.7.1. Asignación de bandas	19
4.8. Quimiometria.....	21
4.8.1. Reconocimiento de pautas no supervisadas.....	23
4.8.1.1. Análisis de agrupamientos	23
4.8.1.2. Análisis de componentes principales	27
4.8.2. Reconocimiento de pautas supervisadas.....	30
4.8.2.1. Modelado independiente de clases: “SIMCA”	32
4.8.3. Modelos de calibración	35
4.8.3.1. Calibración univariante.....	35
4.8.3.2. Regresión lineal múltiple	40
4.8.3.3. Regresión por mínimos cuadrados parciales	41
4.8.3.3.1. Método PLS1	41
4.8.3.3.2. Método PLS2	43
4.8.3.4. Validación de modelos.....	43
5. MATERIALES Y METODOS	47
5.1. Localización del Área de Muestreo.....	47
5.2. Muestreo y Preparación de las Muestras	48
5.2.1. Caracterización físico-química de los suelos	48
5.2.2. Configuración instrumental.....	49
5.2.3. Software para análisis	50
5.3. Experimentación	51

5.3.1. Etapa 1: Análisis de las muestras.....	51
5.3.2 Etapa 2: Generación del modelo de clasificación para identificar suelos agrícolas contaminados con hidrocarburo derivados del petróleo.....	51
5.3.2.1. Análisis Quimiométrico	51
5.3.3. Etapa 3: Análisis soxhlet.....	52
5.3.4. Etapa 4: Estudio de la contaminación de los suelos de la Cuenca Baja del río Tonalá.....	53
6. RESULTADOS Y DISCUSIÓN	55
6.1. Caracterización Físico-Química de los Suelos	55
6.2. Interpretación de los Espectros NIRS de los Suelos.....	57
6.3. Análisis de Componentes Principales (PCA)	60
6.3.1. Análisis de componentes principales de la región espectral NIR.....	63
3.2. Histosol	66
6.3.3. Gleysol.....	68
6.3.4. Solonchack	71
6.3.5. Arenosol.....	73
6.4. Modelado Independiente de Clases (SIMCA).....	76
6.4.1. Modelos de clasificación SIMCA.....	76
6.4.1.1. Modelo de clasificación SIMCA para el Solonchak.....	76
6.4.1.2. Modelo de clasificación SIMCA para el Arenosol.....	80
6.4.1.3. Modelo de clasificación SIMCA para el Gleysol.....	84
6.4.1.4. Modelo de clasificación SIMCA para el Histosol soditerrico	89
6.4.1.5. Modelo de clasificación SIMCA para el Histosol fibrico.....	91
6.4.2. Distancia de clases.....	95

6.4.3. Predicción de clases	96
6.5. Contenido de Hidrocarburos Totales en las Muestras de Suelo	98
6.5.1. Histosol soditérico	99
6.5.2. Histosol fibrico	102
6.5.3. Solonchak.....	102
6.5.4. Gleysol.....	103
6.5.5. Arenosol.....	104
6.6. Modelo para Estimación de Hidrocarburos Totales en Suelos Mediante NIRS	105
6.6.1. Validación interna de las ecuaciones	108
6.6.2. Validación externa del modelo de predicción.....	111
CONCLUSIONES.....	114
LITERATURA CITADA	116
ANEXO	127

Índice de cuadros

Cuadro 1. Número, localización y tipo de emergencia ambiental por derrame de petróleo en México.....	6
Cuadro 2: Límites máximos permisibles de hidrocarburos en suelos (mg kg ⁻¹) de México (NOM-138 SEMARNAT/SS-2003) y en Estados Unidos*	7
Cuadro 3. Regiones del espectro infrarrojo.	15
Cuadro 4. Análisis realizados en las muestras de suelo.	49
Cuadro 5. Valores representativos de las tres primeras componentes principales del espectro UVis-NIR.....	61
Cuadro 6. Valores representativos de las tres primeras componentes principales de los datos del espectro NIR.....	64
Cuadro 7. Valores representativos de los tres primeros componentes principales de los datos del espectro NIR de las muestras del suelo Histosol	66
Cuadro 8. Valores representativos de las tres primeras componentes principales de los datos del espectro NIR de las muestras de suelo Gleysol.	69
Cuadro 9. Valores representativos de las tres primeras componentes principales de los datos del espectro NIRS de las muestras de suelo Solonchack.	71
Cuadro 10. Valores representativos de las tres primeras componentes principales de los datos del espectro NIRS de las muestras de suelo Arenosol.	74
Cuadro 11. Predicción de clases de las muestras de suelo, en función del sitio de muestreo.	98
Cuadro 12. Contenido promedio de hidrocarburos totales en las muestras de suelo.....	101
Cuadro 13. Validación estadística de las ecuaciones de calibración para la validación del modelo de estimación de HTP mediante NIRS.	107
Cuadro 14. Constituyentes estadísticos del modelo de predicción interna.	110

Cuadro 14. Constituyentes estadísticos del modelo de predicción interna. (Continuación).	111
---	-----

Índice de figuras

Figura 1. Principales grupos de hidrocarburos que conforman el petróleo (Dorn <i>et al.</i> , 1998).	9
Figura 2. Técnicas de análisis y separación de hidrocarburos.....	12
Figura 3. Tipos de vibraciones moleculares. Nota: + indica movimiento del plano de la página hacia el lector; - indica movimiento del plano de la página alejándose del lector (Macho, 2002).	18
Figura 4. Frecuencia de vibraciones en el infrarrojo medio (Macho, 2002). .	20
Figura 5. Frecuencia de vibraciones en el infrarrojo cercano (Macho, 2002).	20
Figura 6. Distancia Euclídea y de Manhattan para las muestras k y l.....	24
Figura 7. Unión de un grupo con una muestra aislada de acuerdo con los métodos del centroide (a), de la unión simple (b) y de la unión completa (c).	26
Figura 8. Dendrograma.....	27
Figura 9. Análisis de componentes principales (PCA).....	28
Figura 10. Representación de las puntuaciones de dos clases de objetos en los dos primeros componentes principales.	30
Figura 11. Modelos de dos clases de objetos A y B en el espacio de sus componentes principales, regiones de confianza y predicción de dos muestras	33
Figura 12. Diagrama de Coomans.	34
Figura 13. Matrices de PLS.	42
Figura 14. Espectroscopio NIRs Versatile Agri Analyzer TM 6500	49
Figura 15. Espectro total Nirs de 76 muestras de suelos procedentes de la Cuenca Baja del río Tonalá.	58
Figura 16. Selección región NIR.....	60
Figura 17. Representación de las puntuaciones de las dos primeras componentes principales del espectro total NIR de las muestras de suelo. ..	62

Figura 18. Representación de las puntuaciones de las dos primeras componentes principales del espectro NIR, de las muestras de suelo.....	65
Figura 19. Representación de las puntuaciones de las dos primeras componentes principales del espectro correspondiente a la región NIR de las muestras clasificadas como Histosol.	67
Figura 20. Representación de las puntuaciones de las dos primeras componentes principales del espectro correspondiente a la región NIR de las muestras clasificadas como Gleysol.....	70
Figura 21. Representación de las puntuaciones de las dos primeras componentes principales del espectro correspondiente a la región NIRS de las muestras clasificadas como Solonchack.	72
Figura 22. Representación de las puntuaciones de las dos primeras componentes principales del espectro correspondiente a la región NIR de las muestras clasificadas como Arenosol.	75
Figura 23. Diagrama de Coomans construido con el modelado de clases SIMCA, para las muestras de suelo Solonchack.	78
Figura 24. Diagrama de Coomans construido con el modelado de clases SIMCA para las muestras de suelo Arenosol.....	82
Figura 25. Diagrama de Coomans construido con el modelado de clases SIMCA, para muestras del suelo Gleysol.	86
Figura 26. Diagrama de Coomans, construido con el modelado de clases SIMCA, para muestras del suelo Histosol soditerrico.	90
Figura 27. Diagrama de Coomans construido con el modelo de clases SIMCA, para muestras de suelo Histosol fibrico.	93
Figura 28. Proyección de clases en el modelo SIMCA para las categorías de los cinco suelos.....	97
Figura 29. Modelo de validación interna para estimar TPH mediante NIRS.	108
Figura 30. Modelo de validación externa para estimar TPH mediante NIRS.	113

1. INTRODUCCION

Los hidrocarburos del petróleo en suelos, sedimentos y agua que provienen de derrames de petróleo crudo causan grandes daños a la vida marina y terrestre, a la salud humana y a los recursos naturales. La caracterización adecuada de los hidrocarburos del petróleo crudo derramados y la relación de éstos con las fuentes conocidas son importantes para valorar los daños ambientales. Esta caracterización permitirá entender el destino y comportamiento así como predecir el impacto potencial a largo plazo de los crudos derramados al ambiente, seleccionar las respuestas adecuadas al derrame y tomar medidas efectivas de limpieza (Wang and Fingas, 2003).

La evaluación cualitativa y cuantitativa de los hidrocarburos derivados del petróleo en derrames de petróleo crudo puede realizarse con métodos convencionales o con métodos instrumentales modernos. La utilización de los procedimientos convencionales para resolver la mayoría de los problemas analíticos implica obtener información sobre los compuestos individuales de una mezcla. El tratamiento de la muestra incluye, generalmente, etapas de separación y es la parte más costosa de todo el proceso. Además suele ser, la etapa más larga y es la de mayor fuente de error. El procesado de los datos es sencillo y la información obtenida se utiliza para detectar, identificar y cuantificar los analitos de interés (Guerrero, 2006).

En algunas ocasiones no es necesario disponer de información individual sobre los compuestos de una muestra, cuando el problema no es qué contiene o no contiene la muestra, sino determinar su identidad; esto puede resolverse con señales de perfil del conjunto de compuestos presentes en la muestra. La información contenida en los datos se extrae mediante la

aplicación de técnicas quimiométricas. En este caso la manipulación de la muestra se reduce considerablemente, con lo que disminuyen los errores asociados a esta etapa y el tiempo total de análisis es muy corto, al no ser necesarios los procesos de preparación de las muestras para el análisis. Ambas características ponen de manifiesto que esta estrategia de procesamiento matemático de la información sobre el conjunto de compuestos sea muy atractiva.

El desarrollo de metodologías que permiten la reducción de la etapa de preparación de muestra es, por tanto, una de las tendencias actuales en Química Analítica. El objetivo es obtener la información necesaria para resolver el problema en el menor tiempo posible.

En el presente trabajo se evaluará la capacidad de la espectroscopia del infrarrojo cercano NIRS para generar información útil y suficiente para identificar los suelos agrícolas de la Cuenca Baja del río Tonalá que estén contaminados con hidrocarburos derivados del petróleo crudo; así como para realizar modelos que permitan predecir la concentración del petróleo crudo en los suelos contaminados.

2. OBJETIVOS

2.1. Objetivo General

Desarrollar modelos para clasificación y estimación de la contaminación en suelos agrícolas de la Cuenca Baja del río Tonalá con hidrocarburos derivados del petróleo, mediante la tecnología NIRS.

2.1.1. Objetivos Específicos

2.1.1.1. Generar un modelo de clasificación para identificar suelos agrícolas contaminados con hidrocarburos derivados del petróleo mediante la tecnología NIRS.

2.1.1.2. Generar un modelo de estimación para cuantificar el nivel de contaminación de los suelos agrícolas con hidrocarburos derivados del petróleo mediante la tecnología NIRS.

2.1.1.3 Realizar un diagnóstico de la contaminación por hidrocarburos del petróleo en la Cuenca Baja del río Tonalá con los modelos establecidos.

3. HIPÓTESIS

3.1. Hipótesis General

Los modelos de clasificación y estimación de la contaminación en suelos, generados mediante la tecnología NIRS son suficientes para clasificar y estimar la contaminación por hidrocarburos derivados del petróleo en suelos agrícolas, con resultados similares a los obtenidos con los métodos convencionales.

3.1.1. Hipótesis específicas

Mediante la tecnología NIRS es posible generar modelos que puedan identificar suelos agrícolas contaminados con hidrocarburos derivados del petróleo, similar a los resultados obtenidos con los métodos convencionales.

Mediante la tecnología NIRS es posible generar modelos para estimar los contenidos de hidrocarburos derivados del petróleo en suelos agrícolas, con resultados similares a los obtenidos con los métodos convencionales.

Mediante los modelos generados con el uso de la tecnología NIRS es posible realizar el diagnóstico de la contaminación por hidrocarburos del petróleo en la Cuenca baja del río Tonalá.

4. REVISIÓN BIBLIOGRÁFICA

4.1. La Contaminación con Petróleo Crudo

Las estadísticas mundiales durante el periodo 1978-1997, indican que ocurrieron en promedio 289 derrames de petróleo por año. De ellos, aproximadamente el 74 % de los accidentes corresponden a cantidades entre 40,000 – 400,000 L derramados, 22 % comprendido entre 400,000 y 4,000,000 L, 3 % con 4,000,000 – 40,000,000 L y finalmente 1 % con más de 40,000,000 L (Schmidt, 1999). La tendencia es una disminución de los accidentes (derrames) en un 228 % desde 1970 y en un 154 % desde 1980 (Schmidt, 2001).

La contaminación de suelos con hidrocarburos es un problema que se ha vuelto común en nuestros días, debido al extensivo uso y consumo de combustibles derivados del petróleo en el país. Estos acontecimientos impactan de manera negativa al entorno ecológico (flora y fauna) e incluso la población que está en contacto con los sitios contaminados por hidrocarburos y sustancias químicas (Guerrero, 2006).

En México, la distribución de algunos productos derivados del petróleo se realiza por una red de ductos (70%) que recorre gran parte del territorio nacional, conectando las diferentes estaciones y subestaciones de distribución y venta con las refinerías y plantas petroquímicas. Adicionalmente se utiliza la red de carreteras nacionales (30%) para el transporte de gasolina, diesel y algunos otros derivados del petróleo, lo cual implica un riesgo constante durante el traslado de estas sustancias (Flores *et al.*, 2004) (Cuadro 1).

En 1979, en el Golfo de México ocurrió el mayor derrame de petróleo al mar en el cual tardaron 8 meses en controlar la fuga del pozo petrolero Ixtoc-1 y se derramaron cerca de 700 millones de litros de petróleo en las aguas del golfo.

Cuadro 1. Número, localización y tipo de emergencia ambiental por derrame de petróleo en México.

Año	Número de eventos	Localización				Tipo							
		Terrestre		Marítima		Fuga o derrame		Explosión		Fuego		Otro	
		No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
1993	157	154	98.1	3	1.9	141	89.8	9	5.7	3	1.9	4	2.5
1994	416	389	93.5	27	6.5	359	86.3	21	5.0	28	6.7	8	1.9
1995	547	540	98.7	7	1.3	428	78.2	35	6.4	53	9.7	31	5.7
1996	587	578	98.5	9	1.5	460	78.4	34	5.8	70	11.9	23	3.9
1997	632	574	90.8	58	9.2	541	85.6	49	7.8	26	4.1	16	2.5
1998	538	483	89.8	55	10.2	467	86.8	18	3.3	39	7.2	14	2.6
1999	469	426	90.8	43	9.2	446	95.1	7	1.5	16	3.4	0	0
2000	470	437	93.0	33	7.0	441	93.6	10	2.1	16	3.4	3	0.6
2001	565	530	93.8	35	6.2	517	91.5	17	3.0	19	3.4	12	2.1
2002	470	436	92.8	34	7.2	419	89.1	16	3.4	27	5.7	8	1.7
Total	4851	4547	93.7	304	6.3	4219	87.0	216	4.5	297	6.1	119	2.5

(PROFEPA, 2003).

En el estado de Tabasco, segundo productor de petróleo crudo en México con una producción de 452,859 barriles diarios (INEGI, 2002), en el año 2005 se reportan 25 derrames de hidrocarburos (PROFEPA, 2005). En diagnósticos de suelos realizados en la región, fue posible cuantificar 115,211 mg kg⁻¹ en derrames recientes, y 322,841 mg kg⁻¹ en derrames antiguos (con más de un año), de hidrocarburos totales del petróleo, concentrados en una capa de ≈25 cm de espesor (Rivera-Cruz *et al.*, 2002).

El origen de estos derrames se debe a la rotura de oleoductos, descontrol de pozos, descargas industriales, desbordamiento de presas de descargas, de pozos e instalaciones de almacenamiento, separación y bombeo de hidrocarburos (Botello, 1996; Adams *et al.*, 1996).

4.2. Persistencia y Fraccionamiento de los Hidrocarburos Derivados del Petróleo en el Suelo

Cuando el petróleo llega a la superficie del suelo, por su alta viscosidad no penetra inicialmente hacia el subsuelo, se retiene en el horizonte superficial (PROFEPA, 2000). Una vez presente en el horizonte superficial, se separa en tres fracciones (Kesley *et al.*, 1997):

Volátil: Consiste de la fracción ligera que se fotooxida y volatiliza a través del espacio poroso del suelo y se transporta a la atmósfera, estos son los primeros cuatro n-alcanos (metano, etano, propano y butano).

Disuelta: La fracción soluble que se difunde en la solución del suelo a través del flujo subterráneo de agua. Los hidrocarburos que entran en esta fase son de 5 a 17 carbonos, algunos ejemplos son: pentano, hexano, heptano, benceno, tolueno, etilbenceno, xileno y naftaleno.

Sólida: Es la fracción más estable del petróleo que queda adherida o adsorbida en la matriz del suelo; agrupa a los hidrocarburos con 18 ó más átomos de carbono en su molécula. Incluye a los hidrocarburos aromáticos policíclicos (HAP) y a los compuestos polares.

4.3. Normativas Ambientales

En México se hace uso de la NOM-138-SEMARNAT/SS-2003 "que establece los límites máximos permisibles de hidrocarburos en suelos y especificaciones para su caracterización y restauración" y es muy específica para hidrocarburos del petróleo que provienen de derrames y/o fugas tanto recientes, como pasivos ambientales. En el Cuadro 2 se indican los límites permisibles para los contaminantes altamente tóxicos.

Cuadro 2: Límites máximos permisibles de hidrocarburos en suelos (mg kg^{-1}) de México (NOM-138 SEMARNAT/SS-2003) y en Estados Unidos*

HAP	SUELO			U.S.*
	AGRICOLA	RESIDENCIAL	INDUSTRIAL	
BaA	2	2	10	0.56 – 1.4
DaA	2	2	10	
BbF	2	2	10	0.56 – 1.4
BkF	8	8	80	15 - 5.5
BaP	2	2	10	0.1 - 0.33
Ind	2	2	10	1.5 - 0.56
DaA	2	2	10	0.09 - 0.33

BaA= benzo(a)antraceno, DaA=dibenzo(a)Antraceno;
 BbF=benzo(b)fluoranteno; BkF=benzo(k)fluoranteno; BaP=Benzo(a)pireno;
 Ind=Indenol; DaA=dibenzo(a)antraceno.
 Kostecki *et al.*, 2001.

4.4. Composición Química y Propiedades Físicas del petróleo Crudo

La industria mundial de hidrocarburos líquidos clasifica el petróleo de acuerdo a su densidad establecida como “Parámetro Internacional del Instituto Americano del Petróleo” (API) el cual diferencia los crudos por su calidad (IMP, 2008)

- Ligero. Tiene una densidad en el intervalo de 0.87-0.83 g/cm³ y en grados API, en el intervalo de 31.1-39.
- Pesado. Tiene una densidad en el intervalo de 1.0-0.92 g/cm³ y en grados API, en el intervalo de 10.0-22.3.
- Chapo. Tiene una densidad mayor a 1.0 g/cm³ y de 10.0 grados API.

El petróleo contiene gran variedad de hidrocarburos saturados, y sus productos tienen usos diversos como: gasolina, keroseno, solventes, petróleo crudo y parafinas, entre otros. Estos productos contienen mezclas de hidrocarburos que varían desde líquidos ligeros hasta sólidos. Los hidrocarburos de bajo peso molecular son gases, mientras que los de peso molecular más elevados, son líquidos o sólidos a temperatura ambiente y son insolubles en agua (Dorn *et al.*, 1998).

4.4.1. Hidrocarburos derivados del petróleo

Los hidrocarburos pueden ser saturados e insaturados, de cadena abierta (alifáticos) o cíclicos. En el grupo de los cíclicos, se encuentran los aromáticos, derivados del benceno que presentan resonancia (dobles enlaces no localizados). Los cíclicos no aromáticos presentan dobles o triples enlaces bien definidos y no retienen resonancia. Los enlaces covalentes entre los átomos de carbono de los hidrocarburos, pueden ser simples, dobles o triples (Figura 1).

Estos hidrocarburos tienen enlaces carbono-carbono y carbono-hidrógeno, son polares e insolubles por lo que se disuelven bien en solventes no polares. Debido a su baja polaridad, los hidrocarburos que tienen de uno a cuatro átomos de carbono por molécula, son generalmente gases a temperatura ambiente. Los hidrocarburos que tienen de 5 a 16 átomos, son líquidos a temperatura ambiente. Cuando los alcanos tienen 18 ó más átomos de carbono, son parafinas sólidas a temperatura ambiente.

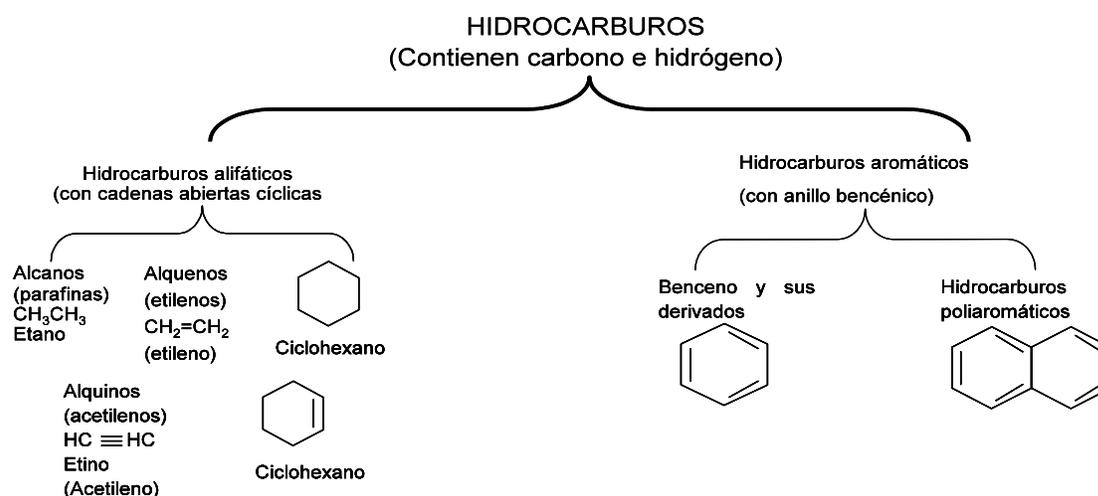


Figura 1. Principales grupos de hidrocarburos que conforman el petróleo (Dorn *et al.*, 1998).

4.4.1.1. Compuestos saturados e insaturados

Una forma sencilla de clasificar a los compuestos orgánicos se basa en la presencia o ausencia de sus enlaces. Si las moléculas de un compuesto incluyen enlaces sencillos se les denomina compuestos saturados o alcanos; son hidrocarburos con moléculas de una sola ligadura (metano, etano, propano). En cambio los compuestos que contienen dobles o triples enlaces, se denominan compuestos insaturados, entre los que se encuentran a los alquenos y alquinos, los alquenos tienen doble ligadura y los alquinos tienen triple ligadura (Dorn *et al.*, 1998).

4.5. Técnicas para la Cuantificación de Hidrocarburos del petróleo en Suelos

4.5.1. Métodos convencionales

Los métodos convencionales para el análisis de hidrocarburos propuestos por la *Environmental Protection Agency* (EPA) (Physical/chemical Methods, 1986), constan habitualmente de tres etapas, extracción, purificación y medida. Destacan la extracción Soxhlet (método 3540C), la extracción en fase sólida (método 3535), la extracción con fluidos presurizados (método 3545), la extracción con ultrasonidos (método 3550B) y la extracción con fluidos supercríticos (método 3560).

Los extractantes más usados para la separación de los hidrocarburos son: hexano, metanol, éter etílico, diclorometano, butanol, acetato de etilo, acetona, ciclohexano o combinación de ellos (Zhang *et al.*, 2002, Szolar *et al.*, 2002). La extracción con Freón 113 (triclorotrifluoretano) fue la mejor opción

y de las más utilizadas pero actualmente su uso ha sido prohibido por reducir la capa de ozono (Lambert *et al.*, 2001).

Además del uso de disolventes es común utilizar tratamientos previos a la extracción con agentes desecantes como el Na_2SO_4 (método 3540C), o la acidificación de las muestras con H_2SO_4 (Erikson *et al.*, 2001).

A fin de agilizar la separación de los analitos de matrices complejas, como los suelos, existe interés por desarrollar técnicas que permitan una extracción eficiente en un tiempo reducido en un alto nivel de automatización (Camel, 2002) y con el empleo de una cantidad pequeña de muestra y disolvente (miniaturización) (Valcarcel y Cárdenas, 2000). Las técnicas para la evaluación de hidrocarburos se dividen en el análisis y en la separación, y se pueden citar a las siguientes (Figura 2).

Técnicas como la interpretación de fotografía aérea, o las imágenes de satélite son también consideradas en los estudios ambientales para la identificación de contaminación y sus fuentes, tanto en mar como en tierra (Perrota y Xefteris, 1996).

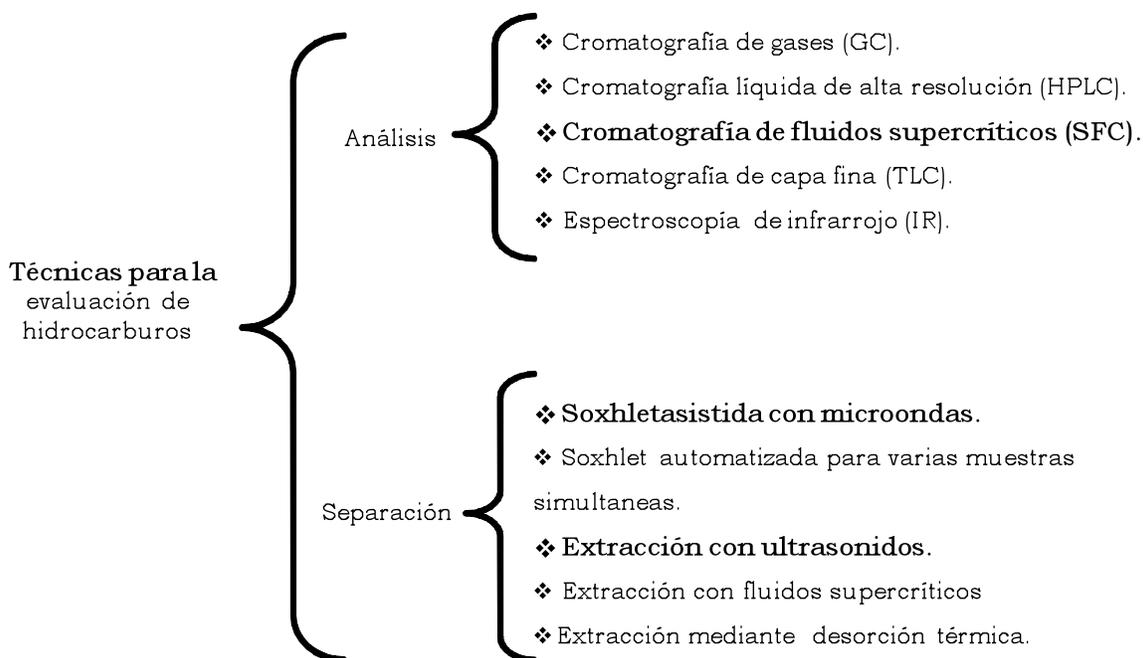


Figura 2. Técnicas de análisis y separación de hidrocarburos (Dean, 1998; Fernández-Linares, 2000; Szolar *et al.*, 2002; Luque-García y Luque, 2001; Guerrero, 2006).

4.5.2. Métodos modernos

Los métodos modernos presentan, actualmente, un gran interés principalmente por su elevada velocidad de análisis. Algunas de las técnicas basadas en este tipo de generación de señales del conjunto de componentes de una muestra son: espectroscopia en el infrarrojo cercano (NIR) (Wang *et al.*, 1995, Macho y Larrechi., 2002, Zhang *et al.*, 2002, González *et al.*, 2003) e infrarrojo medio (MIR) (Carvalho *et al.*, 2006), espectroscopia ultravioleta-visible (Thurston *et al.*, 2004) y pirolisis acoplada a espectrometría de masas (Goodacre y Kell, 1996, Radovic *et al.*, 2001, Pérès *et al.*, 2002), entre otros. También puede incluirse en este grupo de técnicas, a la que suele denominarse olfatometría electrónica.

4.5.2.1. Olfatometría electrónica

La olfatometría electrónica es una metodología de análisis, que responde al esquema de los métodos no separativos, en la que se miden las señales del conjunto de compuestos volátiles de una muestra. Puede ser considerada como una aproximación artificial al sistema olfativo humano diseñada para proporcionar una herramienta rápida y eficaz en el estudio de olores (Guerrero, 2006).

4.5.2.2. Equipos basados en espectrometría de masas

Las etapas del proceso de medida con esta configuración instrumental son idénticas a las ya descritas anteriormente. La respuesta que se obtiene suele estar constituida por un elevado número de picos que corresponden a los fragmentos iónicos generados y que se caracterizan por su relación masa/carga (m/z). La mezcla de compuestos volátiles da lugar a un espectro de masas complejo, con señales formadas por las contribuciones de varios compuestos. De modo análogo a como se trabaja con un conjunto de sensores, las señales obtenidas se tratan mediante técnicas quimiométricas (Guerrero, 2006).

4.5.2.3. Espectroscopía en el infrarrojo cercano (NIRS)

NIRS es una técnica rápida, no destructiva ni contaminante, de gran exactitud siempre que se sigan los procedimientos adecuados para crear modelos de clasificación o de predicción (ecuaciones de calibración). Las propiedades ópticas de los alimentos y en particular la aplicación de la

espectroscopia de reflectancia en el infrarrojo cercano (NIRS) se emplea desde la década del 70 en la industria alimenticia, farmacéutica, petroquímica, entre otras, como alternativa a los métodos químicos y químico-biológicos tradicionales.

La espectroscopia visible (NIR) es una técnica analítica rápida no destructiva, que correlaciona la radiación infrarroja cercana con las propiedades químicas y físicas de los materiales (Chang and Laird, 2002) y se ha utilizado para evaluar las cualidades del suelo (Morra *et al.*, 1991; Confalonieri *et al.*, 2001) y ha demostrado ser rápida, cómoda, sencilla, precisa y capaz de analizar muchos de los elementos constituyentes de las muestras al mismo tiempo.

4.5.3. Aplicaciones de la tecnología NIRS.

El uso de la tecnología de espectroscopia del infrarrojo cercano (NIRS) se ha desarrollado en diferentes campos de la ciencia entre los cuales destacan el área de los alimentos, en estudio de componentes del suelo, así como en la industria farmacéutica.

En la industria se han realizado estudios aplicados en el análisis de alimentos para animales (Cozzolino, 2002), en la predicción de los minerales traza en las legumbres (Cozzolino y Moron, 2004); para determinar la calidad interna de la pera (Liu and Ouyang, 2008); en el control de la calidad interna del tomate (Flores *et al.*, 2009); en la determinación de los contenidos de sacarosa en chocolate (Da costa, 2008).

En estudio del suelo, Velásquez *et al.* (2005) evaluaron la calidad del suelo en los agroecosistemas tropicales de Colombia mediante NIRS, para

discriminar los suelos con diferentes contenidos y características de materia orgánica, sustancias químicas, y biológicas; Cozzolino y Moron (2006), en la predicción de las fracciones de carbono orgánico en el suelo; Chodak *et al.*, 2007, en el análisis de las propiedades químicas y microbiológicas de horizontes orgánicos de los suelos forestales contaminados con metales pesados; He *et al.*, 2007 estudiaron los contenidos de macronutrientes del suelo.

4.6. Teoría de la Absorción en el Infrarrojo

La región infrarroja abarca las regiones del espectro comprendidas entre los números de onda 12800 a 10 cm^{-1} aproximadamente, lo que corresponde a las longitudes de onda 0.78 a 1000 μm^{-1} . Tanto desde el punto de vista de las aplicaciones como de los instrumentos conviene subdividir la región infrarroja del espectro en tres porciones denominadas infrarrojo cercano, medio y lejano; en el Cuadro 3 se indican los límites de cada una de ellas.

Cuadro 3. Regiones del espectro infrarrojo.

Región	Intervalos en longitudes de onda (λ)	Intervalos en números de onda (σ)	Intervalos en frecuencia (ν)
	μm	cm^{-1}	Hz
Cercano	0.78 a 2.5	12800 a 4000	$3.8 \cdot 10^{14}$ a $1.2 \cdot 10^{14}$
Medio	2.5 a 50	4000 a 200	$1.2 \cdot 10^{14}$ a $6.0 \cdot 10^{12}$
Lejano	50 a 1000	200 a 10	$6.0 \cdot 10^{12}$ a $3.0 \cdot 10^{11}$
Más utilizado	2.5 a 15	400 a 670	$1.2 \cdot 10^{14}$ a $2.0 \cdot 10^{13}$

Macho, 2002.

Para absorber radiación infrarroja, una molécula debe experimentar un cambio neto en el momento dipolar como consecuencia de su movimiento vibratorio o rotatorio. Solo en estas circunstancias puede interaccionar con la molécula el campo eléctrico alternante de la radiación y causar cambios en su movimiento. El momento dipolar está determinado por la magnitud de la diferencia de cargas y la distancia entre ambos centros de carga. Si la

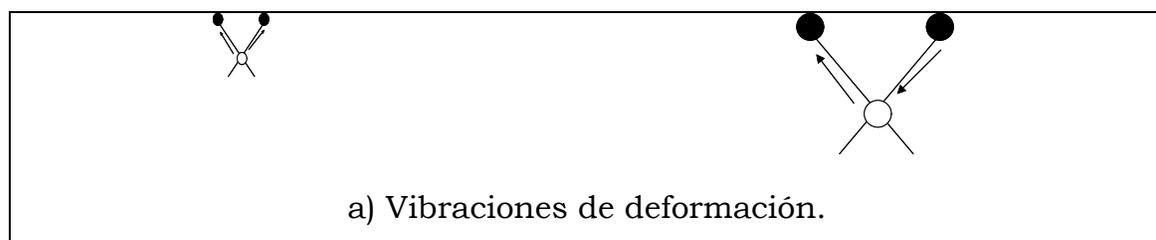
frecuencia de radiación es igual a la frecuencia de una vibración natural de la molécula, ocurre una transferencia neta de energía que da como resultado un cambio en la amplitud de la vibración molecular; la consecuencia es la absorción de la radiación. Análogamente, la rotación de moléculas asimétricas alrededor de sus centros de masa produce una fluctuación dipolar periódica; nuevamente, es posible la interacción con la radiación. No hay cambio en el momento dipolar durante la vibración o rotación de especies homonucleares como O₂, N₂ ó Cl₂ y por tanto estos compuestos no absorben en el infrarrojo (Skoog y West, 1989).

La energía requerida para causar un cambio en el nivel rotatorio es muy pequeña y corresponde a radiación desde 0.01 cm⁻¹ hasta 100 cm⁻¹. Como los niveles rotatorios están cuantizados, la absorción de los gases en esta región del infrarrojo lejano se caracterizan por líneas discretas bien definidas, sin embargo, en líquidos o sólidos los choques intramoleculares e interacciones causan ensanchamiento de las líneas, dando un aspecto continuo. Los niveles de energía vibratoria también están cuantizados, y las diferencias de energías entre estados cuánticos corresponden a las regiones fácilmente accesibles del infrarrojo de 13000 a 675 cm⁻¹ (0.75 a 15 μm). El espectro de un gas consiste, por lo general, en una serie de líneas muy próximas entre sí debido a que existen varios estados energéticos rotatorios para cada estado vibratorio. Por otra parte, en el estado líquido y sólido está muy restringida la rotación, y desaparecen las líneas vibratorias discretas, dejando solo picos vibratorios algo ensanchados. Las posiciones relativas de los átomos en una molécula no están exactamente fijas, sino que fluctúan continuamente como consecuencia de multitud de tipos de vibraciones.

En una molécula diatómica o triatómica simple es fácil definir el número y la naturaleza de tales vibraciones y relacionarlos con las energía de absorción. Sin embargo con moléculas poliatómicas un análisis de esta clase se hace difícil, si no imposible, no solo a causa del gran número de centros

vibratorios, sino también porque ocurren interacciones entre varios centros que deben tomarse en consideración. Las vibraciones corresponden a las categorías básicas de extensión y flexión. Una vibración por extensión supone un cambio continuo en la distancia interatómica a lo largo del eje del enlace entre átomos. Las vibraciones por flexión se caracterizan por un cambio en el ángulo de dos enlaces y son de cuatro tipos: de tijeras, de oscilación, de sacudida y de torsión (Alomar, 1998).

Los distintos tipos de vibraciones se representan en la Figura 3. En una molécula que contiene más de dos átomos, todos los tipos de vibración mostrados pueden ser posibles. Además, puede haber interacción o acoplamiento de vibraciones si éstas suponen enlaces con un átomo central. El resultado del acoplamiento es un cambio en las características de las vibraciones.



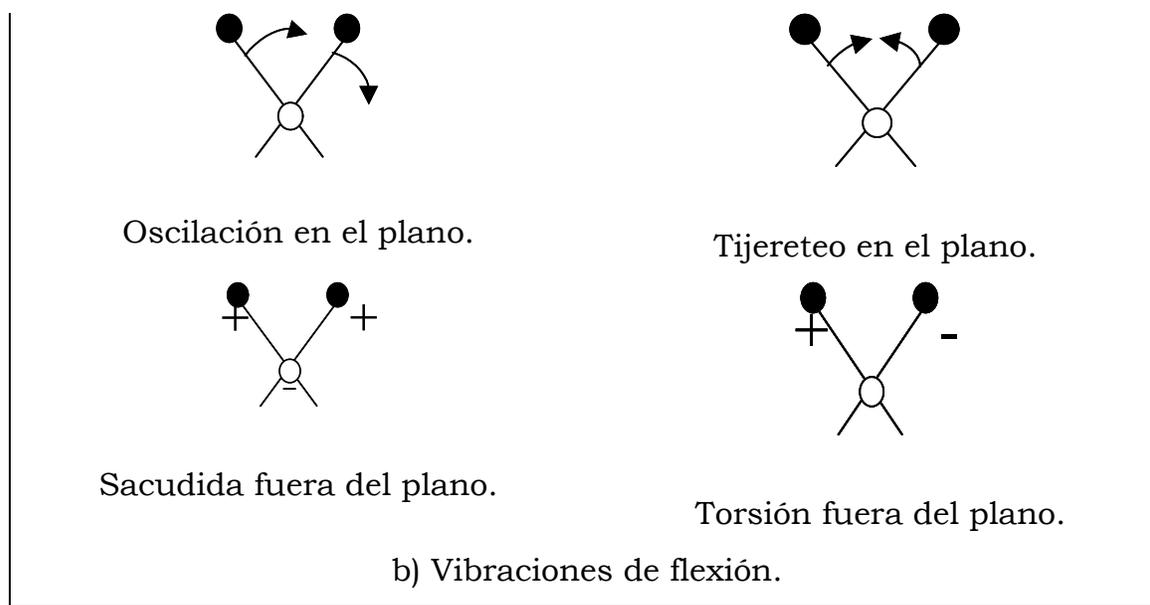


Figura 3. Tipos de vibraciones moleculares. Nota: + indica movimiento del plano de la página hacia el lector; - indica movimiento del plano de la página alejándose del lector (Macho, 2002).

La absorción de radiación infrarroja está confinada para especies moleculares donde existe una diferencia de energía entre los distintos estados rotacionales y vibracionales. La energía requerida para causar un cambio en estados rotacionales es mucho más pequeña que para estados vibracionales y las bandas de absorción rotacional solo se observarían para el caso de los gases. Por lo tanto, para el estudio de espectros de infrarrojo de muestras sólidas y líquidas sólo es necesario considerar los movimientos vibracionales. Las energías vibracionales, como todas las energías moleculares, están cuantificadas y las permitidas para un sistema particular, se encontrarán resolviendo una serie de ecuaciones diferenciales conocidas como ecuaciones de onda mecánico cuánticas (Macho, 2002).

4.7. Interpretación de Espectros

4.7.1. Asignación de bandas

En el espectro infrarrojo medio, entre 4000 y 1300 cm^{-1} o números de onda (región de frecuencias de grupo) se observan una serie de bandas asignadas a vibraciones de sólo dos átomos de la molécula. En este caso la banda de absorción se asocia únicamente a un grupo funcional y a la estructura molecular completa, aunque hay influencias estructurales que provocan desplazamientos significativos en la frecuencia de la vibración. Estas vibraciones derivan de grupos que contienen hidrógeno (C-H, O-H y N-H) o grupos con dobles o triples enlaces aislados. Entre 1300 y 400 cm^{-1} (*fingerprint* región) la asignación a grupos funcionales determinados es más difícil debido a la multiplicidad de bandas, pero es una zona de espectro útil para la identificación de compuestos específicos (Keller *et al.*, 1998). En el espectro de infrarrojo cercano, predominan las bandas debidas a sobretonos y combinaciones de enlaces en los que participa el hidrógeno (debido a que aumenta el grado de anarmonicidad de la vibración), en las Figuras 6 y 7 se muestra un resumen de las absorciones en el infrarrojo cercano (Macho, 2002).

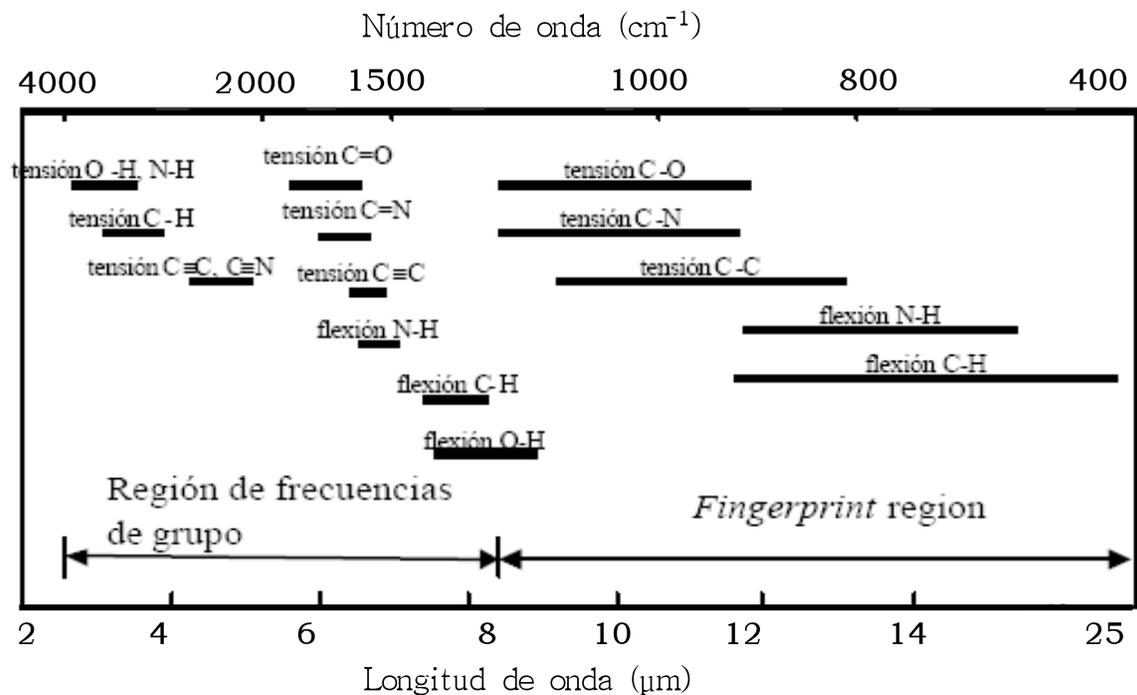


Figura 4. Frecuencia de vibraciones en el infrarrojo medio (Macho, 2002).

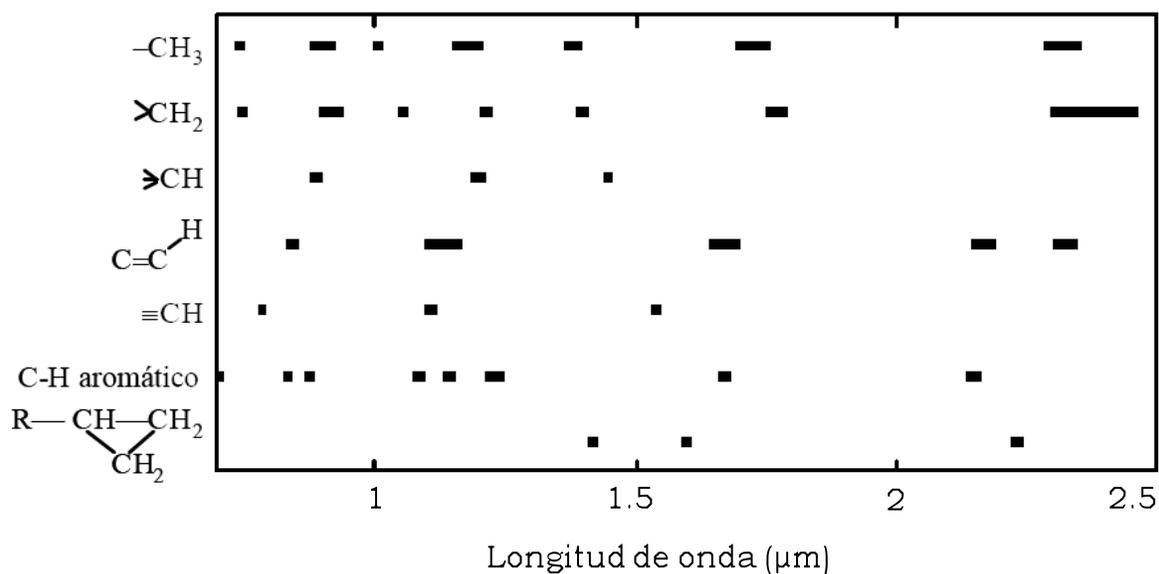


Figura 5. Frecuencia de vibraciones en el infrarrojo cercano (Macho, 2002).

4.8. Quimiometria

La quimiometría puede definirse como una disciplina química que utiliza métodos matemáticos y estadísticos para diseñar procedimientos de medida y para obtener la mayor cantidad de información química posible a partir del análisis de datos (Massart *et al.*, 1997). Trata, específicamente, de todos aquellos procesos que transforman señales analíticas y datos más o menos complejos en información. Utiliza métodos de origen matemático, estadístico y otros procedentes del campo de la lógica formal para conseguir sus fines.

Por todo ello, la quimiometría se sitúa en un campo interdisciplinar. Aunque sus métodos y herramientas provienen de otras disciplinas (como, de hecho, ocurre habitualmente en la química analítica), claramente los fines de la quimiometría están ligados a la química y su éxito depende de los problemas químicos que sea capaz de resolver (Forina, 1993).

Dentro de la quimiometría pueden citarse los diseños experimentales, métodos de exploración de datos, los métodos de reconocimiento de pautas supervisadas, calibración, regresión simple y multivariante así como los sistemas expertos e inteligencia artificial, entre otros.

En el análisis multivariante cada muestra puede considerarse como un punto en un espacio multidimensional. El objetivo es extraer la información útil de todo el conjunto de las variables medidas en una experiencia. El diseño de experimentos es un complemento de este análisis porque proporciona una estrategia para diseñar la experimentación de un proceso. Esto permite obtener la mayor información útil posible con el menor número de experimentos. Además, la calidad de los parámetros calculados depende exclusivamente del diseño seleccionado y ésta es óptima (menor varianza) cuando la matriz del diseño es ortogonal, es decir, las concentraciones de los diferentes compuestos no están correlacionadas (Massart *et al.*, 1997).

Los métodos de reconocimiento de pautas tienen la finalidad de clasificar o agrupar muestras según las características de las mismas (Brereton, 1990). Pueden distinguirse dos tipos de técnicas de reconocimiento de pautas: supervisadas y no supervisadas.

Las técnicas no supervisadas tienen como objetivo determinar relaciones entre muestras sin utilizar ninguna información previa con respecto a la pertenencia a determinadas clases. Permiten conocer la existencia de estructuras en los datos y tener una idea de la utilidad de la información contenida en los mismos. Entre estos métodos se encuentran el análisis de agrupamientos, una de cuyas modalidades es el agrupamiento jerárquico (HCA) en el que las muestras se asocian en función de unos índices de similitud, y el análisis de componentes principales (PCA), en el que se reduce la dimensión de un conjunto de datos mediante el cálculo de combinaciones lineales de las variables originales (Brereton, 2003).

Las técnicas de reconocimiento de pautas supervisadas tienen como fin desarrollar reglas de clasificación para muestras (objetos) desconocidas a partir de un grupo de muestras conocidas que están caracterizadas por los valores de las variables medidas. El éxito de la predicción depende de que las reglas de clasificación y los límites marcados por ellas sean adecuados. El objetivo final es predecir a qué grupo (clase) pertenece una muestra desconocida. Algunas de las técnicas más importantes son el análisis discriminante lineal (LDA), modelado independiente de clases (SIMCA), “support vector machines” (SVM), el método de los k vecinos más próximos (k-NN) y las redes neuronales artificiales (ANN), entre otros (Brereton, 2003).

4.8.1. Reconocimiento de pautas no supervisadas

4.8.1.1. Análisis de agrupamientos

En esta técnica de análisis exploratorio se calcula y compara la similitud entre las muestras. Algunas de las posibilidades para determinar la similitud entre los objetos se describen a continuación (Brereton, 2003).

- Coeficientes de correlación entre las muestras.

Un coeficiente de correlación de 1 implica que las muestras tienen características idénticas. Este coeficiente se obtiene cuando una muestra se compara con ella misma. Los valores pueden oscilar entre -1 y 1 aunque es habitual trabajar con el cuadrado de este número, es decir, el cuadrado del coeficiente de correlación, conocido como coeficiente de determinación.

- Distancia Euclídea.

La distancia Euclídea entre dos muestras k y l está definida por:

$$d_{kl} = \sqrt{\sum_{j=1}^J (x_{kj} - x_{lj})^2} \quad (1)$$

donde J representa el número total de variables medidas y x_{kj} y x_{lj} son los valores de la variable j para las muestras k y l respectivamente. Esta medida es, en cierta forma, contraria al coeficiente de correlación y estrictamente es una medida de “no similitud”. No hay restricciones para su valor si bien es siempre positivo. La ecuación (2) es la notación matricial de la ecuación (1):

$$d_{kl} = \sqrt{(\mathbf{x}_k - \mathbf{x}_l) \cdot (\mathbf{x}_k - \mathbf{x}_l)} \quad (2)$$

- Distancia de Manhattan.

La distancia entre dos muestras k y l se determina según:

$$d_{kl} = \sum_{j=1}^J |x_{kj} - x_{lj}| \quad (3)$$

Las diferencias entre las distancias Euclídea y de Manhattan se ilustran en la Figura 6 en el caso de dos variables. Esta distancia es siempre mayor o al menos igual que la Euclídea

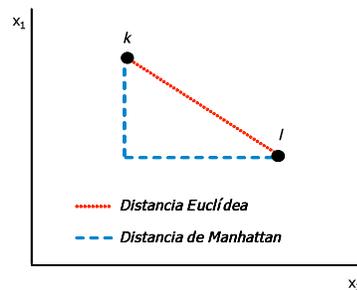


Figura 6. Distancia Euclídea y de Manhattan para las muestras k y l .

- Distancia de Mahalanobis.

Se asemeja a la distancia Euclídea pero tiene en cuenta que algunas variables pueden estar correlacionadas y por lo tanto miden en cierto grado la misma propiedad. La distancia entre dos objetos k y l se puede definir como:

$$d_{kl} = \sqrt{(x_k - \bar{x}_l) \mathbf{C}^{-1} (x_k - \bar{x}_l)'} \quad (4)$$

donde \mathbf{C} es la matriz varianza-covarianza de las variables. Es una matriz simétrica cuyos elementos representan las covarianzas entre pares de variables. Los elementos de la diagonal corresponden a la varianza de cada variable. La matriz es de dimensiones $J \times J$

Este método no es fácilmente aplicable cuando el número de variables excede al de muestras analizadas debido a que la matriz C no tiene inversa. Una posible solución requeriría seleccionar un número más pequeño de variables. Su utilización requiere trabajar con componentes principales en lugar de con las variables originales (ver próxima sección, PCA). El análisis de agrupamientos puede dividirse en dos tipos de técnicas: agrupamientos jerárquicos y no jerárquicos (Bouveresse, 1997).

En los agrupamientos jerárquicos se generan dendogramas en los que se visualizan las distancias entre muestras y en los que pueden considerarse diferentes grupos en función de determinados niveles de similitud.

En los agrupamientos no jerárquicos es necesario prefijar el número de grupos en los que se quieren dividir las muestras y éstos se generan mediante un proceso iterativo en el que la pertenencia de cada muestra a un grupo concreto se va modificando hasta llegar a una situación estable.

En el caso del análisis de agrupamientos jerarquizados (HCA), después de calcular las distancias entre todos los pares de muestras, el siguiente paso es generar los agrupamientos. Los pasos que se deben seguir son los siguientes (Brereton, 2003):

- a) Encontrar las dos muestras más similares, esto es, aquellas entre las que existe la menor distancia.
- b) Formar un grupo con estas muestras. Ahora se sustituye el grupo ya formado por un punto que lo represente. Para seleccionar al representante se suelen aplicar los métodos que se ilustran en la Figura 7 (Ramis, 2001). En el método del centroide, los puntos que ya forman un grupo quedan representados por su centroide. Sin embargo en los otros métodos, el grupo queda representado por aquel punto que presente la menor distancia con respecto a la nueva muestra en el caso de la unión simple y por el de mayor distancia si se aplica el

método de la unión completa. A continuación se recalculan las nuevas distancias entre el nuevo grupo formado y todas las demás muestras buscando de nuevo el valor más pequeño.



Figura 7. Unión de un grupo con una muestra aislada de acuerdo con los métodos del centroide (a), de la unión simple (b) y de la unión completa (c).

- c) Los siguientes pasos consisten en agrupar los datos de manera similar hasta que todos los objetos se unan.

El resultado es una representación gráfica llamada dendrograma. Es una estructura en forma de árbol donde la longitud de las ramas es proporcional a la distancia entre los grupos unidos. La forma del dendrograma depende del método de unión. Habitualmente, las distancias entre muestras se transforman según la siguiente escala de similitud:

$$S = 1 - \frac{d_{kl}}{d_{max}} \quad (5)$$

donde d_{max} es la mayor distancia del conjunto de datos.

En el ejemplo de la Figura 8 los objetos se dividen formando fundamentalmente 2 agrupamientos (A y B). Las muestras 1, 4, 7 y 2 presentan características similares entre sí pero distintas a las del otro grupo, constituido por las muestras 6, 8, 5 y 3.

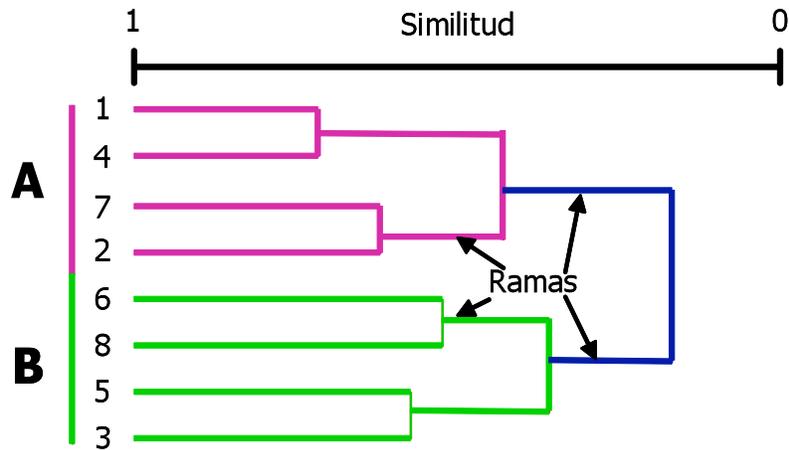


Figura 8. Dendrograma.

4.8.1.2. Análisis de componentes principales

Es una técnica de reconocimiento de pautas no supervisadas que permite encontrar la estructura “latente” en una tabla de datos cuya lectura directa lo impide. El método trata de reducir la dimensión del conjunto de datos construyendo unas nuevas variables, ortogonales entre sí, que reciben el nombre de componentes principales (PCs) y que son combinaciones lineales de las variables primitivas (Brereton, 1992).

En este análisis la matriz de datos \mathbf{X} formada por I objetos o muestras y J variables (dimensión $I \times J$) se descompone (Massart *et al.*, 1997) en el producto de las matrices \mathbf{T} y \mathbf{P} que contienen la información relacionada con las muestras y con las variables respectivamente.

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P} + \mathbf{E} \quad (6)$$

En la expresión (6) la matriz \mathbf{T} contiene las puntuaciones, valores (*scores*) de cada muestra en las nuevas variables, de I objetos en A componentes principales. La matriz \mathbf{P} contiene los pesos, coeficientes (*loadings*) de las variables originales para generar las nuevas, de J variables sobre A

componentes principales. \mathbf{E} es una matriz residual. Las matrices de puntuaciones y pesos consisten en una serie de vectores columna y fila, respectivamente, según muestra la Figura 9. Ambas matrices están formadas por tantos vectores como componentes principales (PCs) puedan calcularse. Estos vectores se representan como t_a y p_a donde \mathbf{a} es el componente considerado y reciben el nombre de vectores principales. Si la matriz de datos originales es de dimensiones $I \times J$, el número máximo de PCs que pueden ser calculados es el valor de la dimensión más pequeña (Massart *et al.*, 1997).

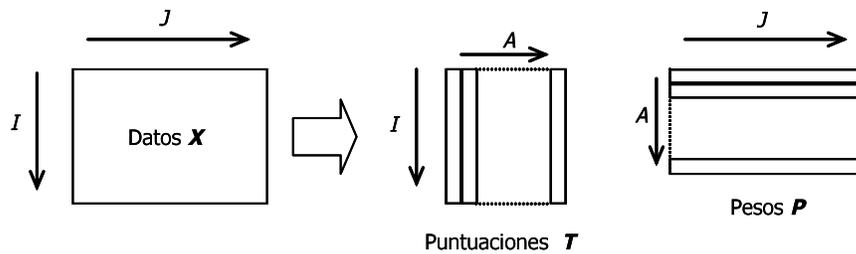


Figura 9. Análisis de componentes principales (PCA).

El objetivo de la técnica es que el número de componentes principales elegido explique la parte determinante de la matriz de datos de modo que los residuales sean lo más pequeños posibles. En el caso de estudiar una mezcla de compuestos (Brereton, 2003), el número de PCs debería ser idéntico al número de compuestos existentes. Sin embargo esto no es generalmente tan simple y debemos tener en cuenta otros factores como por ejemplo el ruido que pueden presentar los datos debido a los errores cometidos en la cuantificación. Los primeros componentes son sin duda los más importantes. El primero explica el mayor porcentaje de varianza de los datos, el segundo el mayor de la restante y así sucesivamente. Mediante el proceso de selección del número de PCs se retiene la información significativa de las matrices y se separa de aquella relacionada con el ruido (Brereton, 2003).

La importancia de cada componente puede medirse mediante la suma de los cuadrados de las puntuaciones según la expresión:

$$g_a = \sum_{i=1}^I t_{ia}^2 \quad (7)$$

donde g_a es el valor propio (*eigenvalue*) del componente principal a . Además la suma de todos los valores de g_a es la suma de los cuadrados de los valores de la matriz de datos \mathbf{X} :

$$\sum_{a=1}^K g_a = \sum_{i=1}^I \sum_{j=1}^J x_{ij}^2 \quad (8)$$

donde K corresponde al número de componentes principales. Frecuentemente estos valores se presentan como porcentajes tal y como muestra la expresión (9).

$$V_a = 100 \cdot \frac{g_a}{\sum_{i=1}^I \sum_{j=1}^J x_{ij}^2} \quad (9)$$

donde V_a es el porcentaje de información modelada por el componente a

El porcentaje acumulado es utilizado a menudo para determinar la proporción de información que ha sido modelada.

Tras aplicar esta técnica, los datos originales quedan matemáticamente modelados por medio de unos pocos componentes principales. Usando un número A de componentes es posible establecer un modelo para cada elemento de \mathbf{X} de la forma:

$$x_{ij} = \sum_{a=1}^A t_{ia} \cdot p_{aj} + e_{ij} = \hat{x}_{ij} + e_{ij} \quad (10)$$

que es la expresión no matricial de la ecuación (6).

Una forma más clara de visualizar esta sección es buscar una interpretación gráfico-geométrica (Blanco y Cerda, 1988) a este análisis. Así, un objeto queda definido por un punto en el hiperespacio constituido por J variables. Esta técnica trata de reducir la dimensión construyendo unas nuevas variables de forma que los objetos estén situados en el nuevo espacio, de forma similar a como lo estaban en el original. Esto produce un giro en el hiperespacio definido por las J variables de forma que la primera variable nueva (primer eje calculado) contenga la máxima información, la segunda (ortogonal a la primera) el máximo de la restante y así sucesivamente. La matriz de puntuaciones está constituida por las proyecciones de cada uno de los objetos del conjunto de datos en los nuevos ejes o componentes principales. La representación de las puntuaciones de los objetos sobre los componentes principales permite reconocer agrupamientos de las muestras; esto se ilustra en la Figura 10.

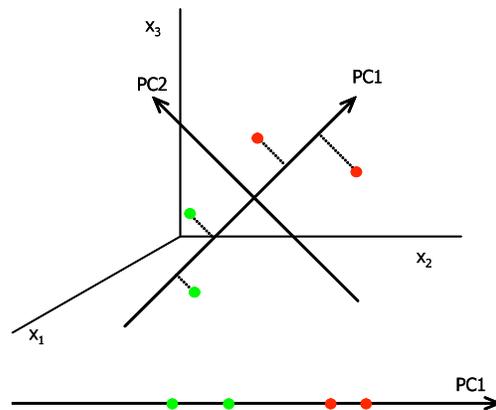


Figura 10. Representación de las puntuaciones de dos clases de objetos en los dos primeros componentes principales.

4.8.2. Reconocimiento de pautas supervisadas

Los métodos de reconocimiento de pautas supervisadas son también conocidos como métodos de clasificación. Su objetivo es clasificar muestras desconocidas en determinados grupos o clases a partir de la información

suministrada por un grupo de entrenamiento. Aunque hay numerosos algoritmos es habitual utilizar una estrategia común para clasificar las muestras independientemente de la herramienta matemática que se utilice. La estrategia consta de tres etapas (Brereton, 2003).

a) Crear un modelo con el grupo de entrenamiento.

El primer paso es generar un modelo matemático que relacione los valores de las variables medidas del conjunto de muestras con el grupo al que pertenecen. Estos objetos reciben el nombre de conjunto de entrenamiento.

b) Comprobar la capacidad predictiva del modelo.

Después de generar el modelo se determina su capacidad predictiva. Un parámetro representativo puede ser el porcentaje de objetos correctamente clasificados.

Esto puede realizarse siguiendo distintos criterios. Dos de los más habituales se describen a continuación:

1. Validación cruzada (*cross validation*).

Es una forma de validación interna del modelo. Una muestra o un conjunto pequeño de ellas se excluyen del conjunto de entrenamiento. Ahora se construye un nuevo modelo sin las muestras excluidas y se comprueba la capacidad de predicción del modelo sobre ellas. El procedimiento se repite hasta que todas las muestras hayan quedado fuera del modelo al menos una vez.

2. Validación externa (*test set*).

Para comprobar la calidad del modelo se realiza la predicción de un grupo independiente de muestras que no ha sido utilizado en ninguna etapa de la construcción del modelo. Se asume que estas muestras son desconocidas y

se comparan los valores obtenidos tras aplicar el modelo con los valores reales conocidos.

Cuando el modelo generado no es satisfactorio existen diferentes posibilidades de mejorarlo. Por un lado se puede utilizar otro algoritmo de cálculo o realizar una selección de variables ya que algunas pueden estar ampliamente influidas por el ruido. En lugar de utilizar todas las variables medidas se seleccionan sólo aquellas que aportan mayor información. El resto podrían estar relacionadas con el ruido.

c) Predicción del conjunto desconocido de muestras.

Una vez que se dispone de un modelo satisfactorio, éste podrá ser aplicado a muestras desconocidas. Un aspecto importante que se debe tener en cuenta es la estabilidad del modelo con el tiempo, es decir, cuando el conjunto desconocido de muestras se analiza semanas o incluso meses después que las correspondientes del grupo de clasificación. En algunas ocasiones es posible que existan variaciones en la respuesta instrumental registrada debido a problemas de inestabilidad de la señal con el transcurso de los días. Esto puede tener seria influencia sobre la capacidad del modelo para clasificar satisfactoriamente una muestra. En problemas de clasificación es habitual realizar algún tipo de normalización de los datos antes de su análisis (Blanco, 1988).

4.8.2.1. Modelado independiente de clases: “SIMCA”

Uno de los métodos de reconocimiento de pautas supervisadas más utilizados es SIMCA (Massart *et al.*, 1997; Brereton, 1990).

Mientras que en otros métodos de clasificación cada muestra se asigna solamente a una clase y ninguna queda fuera de alguna de ellas, éste

permite que el resultado de la clasificación para una muestra sea que no pertenece a ninguna de las clases o que pertenece a más de una (con la misma o diferente probabilidad). En muchas situaciones reales no es posible asignar una muestra concreta a alguna de las clases o existe la posibilidad de que algunas muestras pertenezcan a dos clases simultáneamente.

El método se basa en construir un modelo independiente para cada una de las clases mediante PCA. El número de componentes principales utilizados para cada clase puede prefijarse si se dispone de información previa o mantener un número de ellos que explique un determinado porcentaje de la varianza de los datos. En la Figura 11 se presentan dos clases de objetos, A y B. Al aplicar PCA a la clase A (puntos verdes) se obtiene un modelo de dos componentes (un plano), mientras que la clase B (puntos rojos) se ajusta a un modelo de una única variable (una recta). Las regiones de confianza aparecen alrededor de cada clase (Wold, 1976).

La asignación de los nuevos objetos se hace comprobando si están localizados en la región de confianza de algún modelo. Así el objeto k de la Figura 11 se puede asignar a la clase B mientras que h no pertenece a ninguna clase, es un “outlier”.

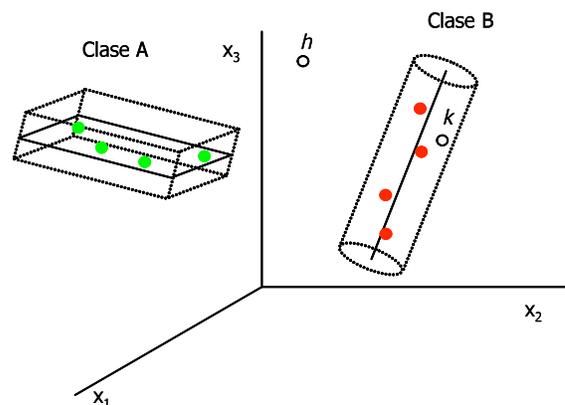


Figura 11. Modelos de dos clases de objetos A y B en el espacio de sus componentes principales, regiones de confianza y predicción de dos muestras.

Para cada clase y su modelo correspondiente se define la sensibilidad como el porcentaje de muestras que perteneciendo a esa clase son reconocidas correctamente por el modelo matemático. Análogamente se define la especificidad como el porcentaje de muestras que perteneciendo a otra clase son reconocidas como ajenas al modelo (Wold, 1976).

Los diagramas de Coomans permiten visualizar estos índices ya que son representaciones bidimensionales de las distancias de las muestras a los modelos. La Figura 12 es un esquema de estos diagramas. En el eje de abscisas se representa la distancia a la clase A y en el de ordenadas a la clase B. La recta horizontal discontinua delimita el recinto de la clase B mientras que la línea vertical discontinua delimita el recinto de la clase A. El cuadrado inferior izquierdo (1) es el recinto común a ambas clases y cuanto menor sea el número de muestras en sus coordenadas, mayor será la especificidad de los modelos. El cuadrado superior derecho (2) representa el espacio que queda fuera de ambos modelos. Las muestras con coordenadas en él son muestras extrañas a los modelos generados (Coomans y Broeckaert, 1986).

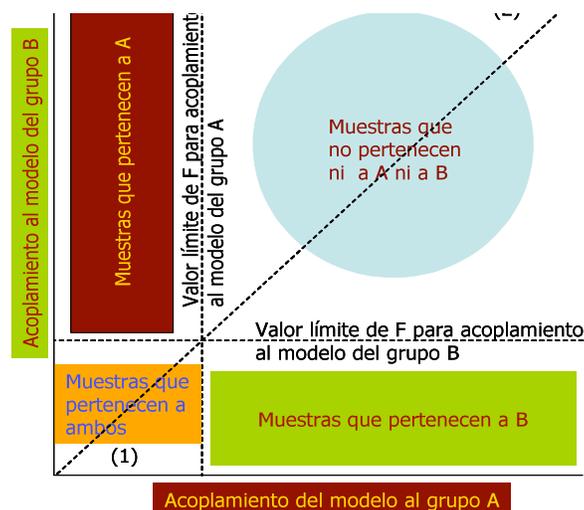


Figura 12. Diagrama de Coomans.

Cuando SIMCA se utiliza únicamente como técnica de clasificación, todas las muestras que se encuentren por encima de la diagonal serán asignadas a la clase A y las muestras por debajo a la clase B.

4.8.3. Modelos de calibración

4.8.3.1. Calibración univariante

La calibración univariante (Massart *et al.*, 1997; Brereton, 2000; Miller and Miller., 2000) permite obtener modelos capaces de predecir una variable respuesta conociendo los valores de otra correlacionada con ella

La expresión característica (Brereton, 2000) de este tipo de calibración es:

$$=Cs..... (13)$$

Donde **X** es un vector que está formado por las respuestas analíticas para el número de muestras analizadas y **c** es el vector que corresponde a las concentraciones. Ambos vectores tienen la misma longitud, *I*, que es igual al número de muestras del escalar S que relaciona ambos parámetros.

En muchas situaciones será apropiado incluir un término adicional en el modelo de calibración que es la ordenada en el origen de la forma:

$$=a+cs (14)$$

La calidad de las predicciones puede determinarse mediante la expresión:

$$E = \sqrt{\frac{\sum_{i=1}^I (c_i - \hat{c}_i)^2}{d}} (15)$$

donde $(c_i - \hat{c}_i)$ es la diferencia entre las concentraciones añadida y predicha para la muestra i . “ d ” son los grados de libertad. En el caso de la calibración univariante el número de los grados de libertad es igual al número de determinaciones (I) realizados menos el número de parámetros calculados en el modelo. Este error absoluto puede representarse como porcentaje mediante la expresión:

$$E\% = \frac{E}{\bar{C}} \cdot 100 \quad (16)$$

Donde \bar{c} es la media de las concentraciones añadidas de todas las muestras.

A partir de los modelos clásicos que incluyen ordenada en el origen es posible determinar tres varianzas de interés (Ramis, 2001).

- La varianza explicada por el modelo.

Es la suma de cuadrados (SC) de las desviaciones de los valores predichos respecto a las coordenadas del centroide dividida por el número de grados de libertad que es uno.

$$S_{\text{exp}}^2 = \sum_{i=1}^I (\hat{c}_i - \bar{c})^2 \quad (17)$$

- La varianza residual.

Es la suma de cuadrados de las desviaciones de los valores reales respecto a los predichos divididos entre el número de grados de libertad, que en este caso es el número de experimentos realizados menos uno. Este tipo de varianza es una medida de la dispersión de las medidas a lo largo de la línea de regresión ajustada y representa la varianza en la respuesta que el modelo no tiene en cuenta.

$$S_{res}^2 = \frac{\sum_{i=1}^I (c_i - \hat{c}_i)^2}{I - 1} \quad (18)$$

➤ La varianza total.

La varianza total de los valores reales en ausencia de modelo se calcula mediante la suma de cuadrados de las desviaciones de los valores reales respecto al centroide dividida por el número de grados de libertad, $I - 1$.

$$S_{tot}^2 = \frac{\sum_{i=1}^I (c_i - \bar{c})^2}{I - 1} \quad (19)$$

De acuerdo con Brereton (2003) la relación entre las tres sumas de cuadrados es:

$$S_{tot}^2 = S_{exp}^2 + S_{res}^2 \quad (20)$$

Lo que manifiesta que el modelo divide la varianza total en dos, la explicada por él, y la residual. Fácilmente se puede demostrar que la relación entre S_{exp}^2 y S_{tot}^2 equivale al cuadrado del coeficiente de correlación lineal (Ramis, 2001):

$$r^2 = \frac{\sum_{i=1}^I (\hat{x}_i - \bar{x})^2}{\sum_{i=1}^I (x_i - \bar{x})^2} = \frac{S_{exp}^2}{S_{tot}^2} \quad (21)$$

El parámetro r^2 se denomina coeficiente de determinación y expresa el porcentaje de variación total de los datos que es explicado por el modelo. Se utiliza para comprobar el porcentaje de información modelada por la regresión pero no necesariamente implica que el modelo más adecuado para ajustar los datos sea el lineal incluso cuando sus valores sean próximos a 1.

Además de obtener un buen coeficiente de correlación es importante comprobar que el modelo generado no contiene fallo de ajuste (Analytical Methods Committee, 1994). Para ello es necesario tener réplicas de cada nivel de concentración.

La secuencia que se representa a continuación corresponde a un conjunto de muestras con n niveles de concentración diferentes, repetidos cada uno, p veces. X_{ij} son los valores experimentales obtenidos para las muestras.

C_1 $X_{11} \dots X_{1j} \dots X_{1p}$

C_2 $X_{21} \dots X_{2j} \dots X_{2p}$

.

C_i $X_{i1} \dots X_{ij} \dots X_{ip}$

.

C_n $X_{n1} \dots X_{nj} \dots X_{np}$

Ahora es posible dividir la variación residual, S_{res}^2 , en dos partes:

➤ Variación debido al error puro.

Mide la variación debida al error experimental según la expresión:

$$S_{error}^2 = \sum_{ij} (x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - x_i)^2 \quad (21)$$

Donde \bar{x}_i es la media de las medidas realizadas para cada nivel de concentración.

- Variación debida al cambio de ajuste

$$\bar{x}_i = \frac{1}{p} \sum_{j=1}^p x_{ij} \quad (22)$$

Es un componente que mide la variación de la media de las muestras repetidas para cada nivel de concentración a lo largo de la línea de regresión mediante la expresión.

$$S^2_{f\ allaj\ u} = S^2_{res} - S^2_{error} \quad (23)$$

Las varianzas asociadas a estos factores se obtienen dividiendo las sumas de cuadrados entre el número de grados de libertad tal y como se indica en las siguientes expresiones:

$$S_{error} = \frac{SC_{error}}{n(p-1)} \quad (24)$$

$$S^2_{f\ allaj\ us} = \frac{SC_{res} - SC_{error}}{n-2} \quad (25)$$

El posible fallo de ajuste se puede comprobar comparando la relación

$$F = \frac{S^2_{f\ allaj\ us}}{S_{error}^2}$$

obtenida experimentalmente con el valor crítico tabulado de F a

(n-2) y n (p-1) grados de libertad y un determinado nivel de significación. Si la diferencia es significativa (la varianza del fallo de ajuste es mayor que la varianza debida al error puro) se puede concluir que el modelo es inadecuado debido a que la variación de las medias correspondientes a las muestras repetidas con el mismo nivel de concentración no puede ser explicado en términos de la incertidumbre del error puro. En este caso es aconsejable adoptar un modelo diferente. En caso contrario el modelo no representa fallo de ajustes y es adecuado para representar el conjunto de muestras estudiadas.

4.8.3.2. Regresión lineal múltiple

La regresión lineal múltiple (MLR) es una extensión del modelo anterior, en la que se utiliza más de una variable predictora o respuesta instrumental. Las principales razones para utilizar estos nuevos métodos son dos en primer lugar es posible que exista más de un compuesto en las muestras que se quieran analizar. En estas circunstancias es necesario utilizar más de una respuesta (Massart *et al.*, 1988).

Para n compuestos al menos deben ser medidas n respuestas diferentes a no ser que se conozca que la concentración de algunos de ellos está correlacionada. La segunda razón está relacionada con el hecho de que cada respuesta contiene algún tipo de información y, aunque algunas pueden contener ruido o intereferentes desconocidos, los modelos que incluyen más de una variable predictora proporcionan generalmente mejores resultados que aquellos contruidos con una sola (Ramis, 2001).

Esta técnica proporciona buenos resultados cuando se incluyen en el modelo todos los compuestos que contribuyen a la señal medida. En caso contrario las predicciones empeoran notablemente. Supongamos una mezcla de diez compuestos. La regresión lineal múltiple proporciona resultados adecuados si se dispone de información suficiente de los diez compuestos. Sin embargo si solo se sabe de la presencia de tres de ellos, las predicciones empeorarán ya que el resto de los compuestos distorsionan el modelo. Los compuestos no modelados contribuyen al cálculo de los coeficientes de las variables predictoras aunque luego se afirma que el modelo formado corresponde a un número menor de compuestos. Este problema podría solucionarse (Brereton, 2000), en parte, si se seleccionan exclusivamente las variables más importantes de los tres compuestos que se estudian. La necesidad de modelar todos los compuestos significativos es la mayor limitación de esta regresión.

4.8.3.3. Regresión por mínimos cuadrados parciales

La regresión por mínimos cuadrados parciales es a menudo considerada como la técnica de regresión multivariante (Martens, 1996) más importante y a diferencia de la técnica anterior no requiere un modelo exacto para todos los compuestos.

El método está basado en análisis de componentes principales y no requiere de detalles de todos los compuestos presentes en las muestras aunque será necesario realizar una estimación del número de componentes principales significativos que no se corresponde necesariamente con el número de especies químicas (Brereton, 2000)

Uno de los aspectos más importantes de esta técnica de calibración frente a otras como la regresión por componentes principales (PCR) es que tiene en cuenta no solo los posibles errores de bloque de datos **X**, sino también los errores relacionados con la estimación de la concentración.

Existen fundamentalmente dos métodos de la regresión por mínimos cuadrados parciales que se denominan PLS1 y PLS2. A continuación se describen ambos.

4.8.3.3.1. Método PLS1

Los bloques de matrices X y c se descomponen en sus componentes principales de acuerdo con las siguientes expresiones (Brereton, 2000):

$$x = T \cdot P + E \quad (26)$$

$$c = T \cdot q + f \quad (27)$$

Estas matrices se han representado en la Figura 13. El producto de las matrices **TP** y **Tq** se aproxima a los valores medidos con la configuración instrumental utilizada y a las concentraciones reales de cada muestra, respectivamente. **E** y **f** son las matrices de error. La matriz de puntuaciones **T** es común a ambos bloques, **X** y **c**.

Como se observa en la Figura 13 para cada componente “a” existe un vector de puntuaciones o “scores”, un vector de pesos o “loadings” y una matriz de residuales (Brereton, 2000).

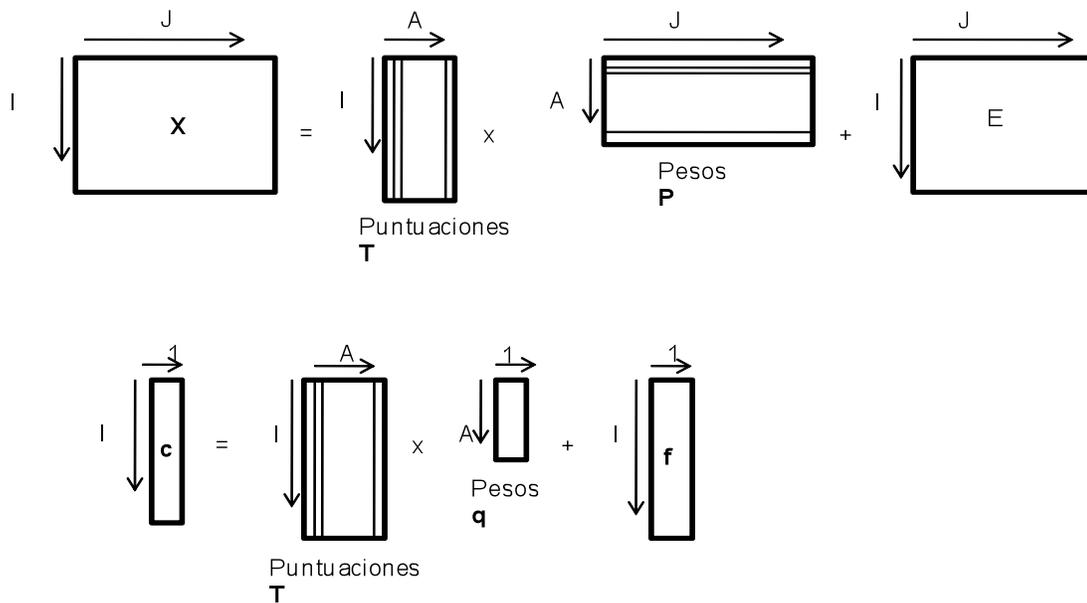


Figura 13. Matrices de PLS.

Cada compuesto considerado genera una matriz de puntuaciones distintas, al contrario que en otras técnicas como PCR donde hay una única matriz de puntuaciones para todos los compuestos de la mezcla.

4.8.3.3.2. Método PLS2

La calibración mediante PLS2 permite el uso de una matriz de concentraciones \mathbf{C} , en lugar de los vectores concentración \mathbf{c} de cada compuesto individual de la mezcla. La ecuación (26) se transforma ligeramente y \mathbf{q} se convierte en la matriz \mathbf{Q} . el número de columnas en \mathbf{C} y en \mathbf{Q} es igual al número de compuestos estudiados. En PLS1 se requiere que cada compuesto sea modelado independientemente mientras que en PLS2 todos se modelan conjuntamente. Las concentraciones estimadas mediante ambas técnicas pueden no ser iguales.

PLS2 puede ser útil (Brereton, 2003), en los casos donde existe correlación entre la concentración de los compuestos estudiados. En lugar de obtener un modelo independiente para cada compuesto de una mezcla, PLS2 permite analizar todos los datos de una vez. Sin embargo en muchas situaciones las predicciones realizadas mediante esta técnica son peores que las estimaciones correspondientes mediante PLS1. Una posible estrategia puede ser realizada por PLS2 sobre el conjunto de datos como primera aproximación y después realizar PLS1 individualmente sobre cada compuesto conocido.

4.8.3.4. Validación de modelos

Uno de los aspectos más importantes de los métodos de calibración es la validación de los modelos. Esta facilita la información fundamentalmente sobre el número de componentes significativos necesarios para caracterizar un conjunto de datos y sobre la capacidad predictiva de los modelos en futuras muestras. Es posible obtener un buen ajuste entre los valores reales y los calculados mediante PLS aumentando el número de componentes considerados. Éste puede llegar, incluso, a ser completo si se utilizan todos

los componentes PLS. Los principales métodos de validación se describen a continuación.

➤ Auto predicción.

Es la aproximación más sencilla y se basa en calcular la raíz cuadrada de a media del error de calibración al cuadrado para cada compuesto, se calcula mediante la expresión:

$$E_{cal} = \sqrt{\frac{\sum_{i=1}^I (c_i - \hat{c}_i)^2}{I}} \quad (28)$$

Donde I es el número de muestras y la expresión entre paréntesis representa la diferencia entre la concentración puesta y la predicha por el modelo. Generalmente este modelo de validación no se utiliza, ya que los datos pueden llegar a tener un ajuste perfecto aumentando simplemente el número de componentes del modelo.

➤ Validación cruzada (*cross validation*).

La capacidad predictiva del modelo se comprueba separando un pequeño número de muestras del conjunto de calibración y analizando las predicciones obtenidas para este conjunto para el modelo reducido.

Aunque existen diversos criterios, es habitual retirar solo una muestra del modelo y predecir su concentración. Esto se repite I veces de modo que al final todas las muestras se han predicho una vez. El error cometido cuando se utiliza la validación cruzada (CV, siglas en inglés) se calcula mediante la expresión:

$$E_{cv} = \sqrt{\frac{\sum_{i=1}^I (C_i - \hat{C}^{cv})^2}{I}} \quad (29)$$

donde \hat{C}^{cv} es la concentración predicha de la muestra i del grupo de validación cruzada.

➤ Validación externa (*test set*).

Uno de los puntos débiles de la validación cruzada es que depende de cómo se ha realizado el diseño experimental para formar el modelo de calibración. Si dos compuestos están correlacionados, de forma que cuando la concentración de uno aumenta la del otro también lo hace y al revés; la auto predicción e incluso la validación cruzada puede sugerir que el modelo de la calibración es adecuado. Sin embargo si una futura muestra de predicción con un contenido elevado de los compuestos presenta un contenido bajo del otro es muy probable que los resultados de predicción se alejen de los valores reales. La validación cruzada es muy útil para informar sobre el ruido instrumental o errores de dilución y pesada por ejemplo, pero no es de gran ayuda si existe correlación en las concentraciones del grupo de patrones de calibración.

En lugar de la validación cruzada es posible comprobar la calidad predictiva del modelo con un conjunto independiente de muestra que no se han utilizado en ninguna etapa durante la construcción del modelo. El error cometido cuando se utiliza un grupo independiente o “*test set*” de I muestras es:

$$E_{test} = \sqrt{\frac{\sum_{i=1}^I (C_i - \hat{C}^{test})^2}{I}} \quad (30)$$

donde $c_i^{test \wedge}$ es la concentración predicha de la muestra i del grupo de validación externa.

A partir de los valores absolutos de las ecuaciones (27-28) es habitual expresar el resultado en términos relativos, nuevamente, mediante la expresión (15).

Finalmente el sesgo (bias) del modelo y del grupo de validación se cuantifica mediante la expresión:

$$S = \frac{\sum_{i=1}^I (C_i - \hat{C})}{I} \quad (31)$$

Donde I es el número de muestras del modelo o del grupo de validación.

5. MATERIALES Y METODOS

El desarrollo experimental se realizó en cuatro etapas, una vez que fueron obtenidas las huellas espectrales mediante la tecnología NIRS de las muestras procedentes de la Cuenca Baja del río Tonalá.

La primera consistió en generar un modelo de clasificación para identificar suelos agrícolas contaminados con hidrocarburos. En la segunda etapa se generó un modelo de estimación para cuantificar el nivel de contaminación de los suelos agrícolas con hidrocarburos derivados del petróleo. Para validar lo anterior, las muestras de suelo fueron analizadas mediante extracción Soxhlet y gravimetría; como método de referencia. Finalmente en la tercera etapa se realizó un diagnóstico de la contaminación con hidrocarburos derivados del petróleo, en los suelos agrícolas de la cuenca baja del río Tonalá.

5.1. Localización del Área de Muestreo

El área de muestro se localiza entre las coordenadas 18° 05' 25" y 18° 18' 52" N y 93° 47' 02" y 94° 06' 23" O, abarca 22,432 ha, en la planicie de inundación del río Tonalá, en el límite oeste de Tabasco. Incluye parte de los municipios de Cárdenas y Huimanguillo, las poblaciones más importantes son La Venta y Benito Juárez y corresponde a la zona de influencia de los campos petroleros Cinco Presidentes, Rodador y La Venta Norte (Anexo, Cuadro 1).

Los suelos están representados por los grupos: Arenosol, Histosol, Gleysol, Solonchak, (Palma-López *et al.*, 1999); su vegetación natural está compuesta por palmar, matorral y selva baja inundable, manglar e hidrófitas, cuenta

además con áreas de pastos inducidos tolerantes a la humedad, cultivos múltiples y perennes (plantaciones de coco). Se localizan 437 pozos petroleros, siete baterías de separación, tres estaciones de compresoras y una planta de inyección de agua, además de una densa red de oleoductos y terracerías.

Estos suelos fueron nuevamente clasificados por Palma-López *et al.*, 2009 y cambiaron sus claves a Arenosol haplico (distrito), Histosol Rheico Saprico (eutrico sodico), Histosol rheico hemico (Districo), Solonchak Gleyico Histico (Sodico) y Gleysol Molico (eutrico sodico) respectivamente (Anexo, Figuras 1, 2, 3 y 4 respectivamente).

5.2. Muestreo y Preparación de las Muestras

El muestreo de los suelos consistió en abarcar la zona de impacto desde un punto que tuviera un grado de contaminación alto, hasta la desembocadura de la cuenca, muestreando en zig-zag y con repeticiones para hacer una muestra compuesta, se colectaron un total de 76 muestras compuestas (Anexo, Cuadro 2) las cuales fueron llevadas a laboratorio en donde se aplicó el proceso de secado bajo sombra a temperatura ambiente y posteriormente se molieron y tamizaron a un diámetro de 2 mm, para realizar los análisis físico-químicos conforme lo establece la NOM-021-SEMARNAT-2000 y las mediciones por espectroscopia en el infrarrojo cercano.

5.2.1. Caracterización físico-química de los suelos

Se realizó la caracterización físico-química de cada uno de los suelos secando las muestras a temperatura ambiente y bajo sombra posteriormente

se analizaron con los métodos establecidos en la Norma Oficial Mexicana, NOM-021-SEMARNAT-2000 (Cuadro 4).

Cuadro 4. Análisis realizados en las muestras de suelo.

Determinación	Tipo de análisis.	NOM-021-RECNAT-2000
pH	Relación 1:2 con agua.	AS-02
Conductividad eléctrica	Pasta de saturación	AS-18
Materia orgánica.	Digestión húmeda.	AS-07
Nitrógeno total.	Kjeldahl	AS-25
Bases intercambiables y cationes intercambiables	Con acetato de amonio	AS-12
Textura.	Bouyoucos.	AS-09

5.2.2. Configuración instrumental

El equipo que se utilizó es un Espectroscopio de Reflectancia en el Infrarrojo Cercano modelo Versatile Agri Analyzer™6500 System II Vertical Type 461008, Serial No. 9503 100-240 Vac max 200 W, 50-60 Hz de la empresa FOSS NIRSystems (Figura 14), el intervalo que cubre el escáner es de 400 a 2500 nm.



Figura 14. Espectroscopio NIRs Versatile Agri Analyzer TM 6500.

Para realizar la lectura el equipo cuenta con una lámpara de 4 Volts, ajustada a una placa prealineada, el cual envía el haz de luz blanca. Ésta contiene todas las frecuencias al *Grating* (que es una superficie holográfica de 570 líneas por mm y una distancia de 1.75 μm entre líneas éste desvía la luz en ángulos diferentes dependiendo de la longitud de onda, ésta luz reflejada es monocromada y su ángulo de inclinación es controlado por la tarjeta madre); al filtro, *Order sorter*, que está sincronizado con el movimiento del *Grating* y es movido por medio de un motor de pasos, permite el paso de la luz por longitud de onda para transmitirla a la muestra, misma que es contenida en una copa o celda opaca de cuarzo. La luz es enviada a los detectores de silicio (400 a 1100 nm) y PbS (1100 a 2500 nm), que registrarán la cantidad de luz reflejada y/o transmitida, para transformarla en una señal que se envía mediante el programa ISIScan (versión 2.81 Copyright 2006) a una computadora (Acer AMD Sempron™ Processor 3200+ 1.81 GHz, 448 MB de RAM).

5.2.3. Software para análisis

La obtención de los espectros de las muestras se realizó con el programa ISISCAN. Para la interpretación quimiométrica de los espectros se usaron como programas puentes al WinISI III y Excel, y para la realización del análisis de componentes principales y el modelado independiente de clases se usó el Pirouette Chemometrics Software versión 3.11 para Windows de Infometrix Inc. Copyright 2007. La prueba de medias de Tukey se realizó con el programa estadístico SAS versión 9.

5.3. Experimentación

5.3.1. Etapa 1: Análisis de las muestras

El análisis de las muestras se realizó colocando una cantidad de muestra en la celda porta muestra con medidas de 10 x 6 cm y fabricada de cuarzo. La colocación de la muestra debe ser compacta, de tal forma que se encuentre homogénea en la celda sin que se le formen grietas.

Las longitudes de onda que se registran durante el análisis van desde 400 hasta 2500 nm. El espectro se divide en dos partes, la primera pertenece a las longitudes de onda de la luz visible, que registra longitudes de onda de 400 hasta 750 nm, y lo que resta del espectro pertenece al infrarrojo cercano.

5.3.2 Etapa 2: Generación del modelo de clasificación para identificar suelos agrícolas contaminados con hidrocarburo derivados del petróleo

La generación de modelos para identificar suelos contaminados con la tecnología NIRS se estudió con una serie de suelos con características fisicoquímicas diferentes (Gleysol, Histosol, Histosol fibrico, Arenosol y Solonchak) los cuales fueron colectados en el sitio descrito en el apartado 5.1.

5.3.2.1. Análisis Quimiométrico

Las señales espectrales obtenidas de las muestras de suelo fueron analizadas quimiométricamente, utilizando el programa Piroutte Chemometrics Software versión 3.11. Para ello se aplicaron las técnicas de

reconocimiento no supervisadas para determinar relaciones entre muestras sin utilizar ninguna información previa con respecto a la pertenencia a determinadas clases. Se aplicó el análisis de componentes principales (PCA) que permitió encontrar la estructura “latente” en la matriz de datos cuya interpretación directa lo impide.

Para la realización de estos análisis quimiométricos a los datos se les realizó un pre-procesado entre los cuales se encuentra: la media central, varianza escalar, autoescala, rango de escala. Posteriormente se les aplicó una transformación a los datos con algunas de las siguientes opciones: primera y segunda derivada, corrección de la línea de la base, dividida por (variables), log 10, normalización, suavizados y la varianza normal estándar entre otros. Estos pre-procesados y transformación se realizarán de acuerdo como se comporte el agrupamiento de las muestras.

Así mismo se aplicó la técnica de reconocimiento supervisada SIMCA, para construir un modelo independiente para cada una de las clases mediante PCA. Y se estableció la frontera entre clases con un nivel de significancia del 5%.

5.3.3. Etapa 3: Análisis soxhlet

El análisis de los contenidos de los hidrocarburos totales del petróleo (TPH) en suelos, se realizó mediante extracción por reflujo Soxhlet establecido por la NOM-138-SEMARNAT/SS-2003, éste método asegura el contacto íntimo de la matriz de la muestra con el solvente de extracción. La extracción de hidrocarburos del petróleo por Soxhlet provee fracciones de C6 a C50. Para la óptima extracción de los compuestos orgánicos, los sólidos deben estar en partículas pequeñas; mientras más pequeñas sean las partículas, más área de superficie y contacto y, por lo tanto, mejor extracción. Se basa en la

cuantificación de los hidrocarburos extraídos utilizando un solvente adecuado. El solvente usado fue el diclorometano el cual es evaporado y el extracto orgánico obtenido es cuantificado gravimétricamente. El resultado se reporta como porcentaje de la muestra total de peso seco (ss).

Los datos obtenidos por el método de extracción Soxhlet, el cual fue usado como un método de referencia, permitieron comprobar si la tecnología NIRS con apoyo de la quimiometria es capaz de clasificar las muestras de suelo en función del grado de contaminación que presentan. Para ello se realizó una prueba de medias de Tukey $P \leq 0.05$ con el programa estadístico SAS.

5.3.4. Etapa 4: Estudio de la contaminación de los suelos de la Cuenca Baja del río Tonalá

El estudio de la contaminación de los suelos de la Cuenca Baja del río Tonalá se realizó simultáneamente en las etapas 5.3.1, 5.3.2 y 5.3.3

En la primera etapa mediante los datos obtenidos por NIRS se llevó a cabo la separación de las muestras de manera cualitativa dependiendo del tipo de matriz de suelo, y por sitio de muestreo de la zona de estudio. Mientras que en la segunda etapa se cuantificó el nivel de la concentración de los hidrocarburos totales del petróleo en suelos contaminados, para así generar modelos que permitan estimar la concentración de los contaminantes del petróleo en suelos.

Con la información en conjunto se realizó el estudio de la contaminación de los suelos de la Cuenca Baja del río Tonalá, mediante espectroscopia del infrarrojo cercano y el método Soxhlet como referencia, y clasificando los niveles de contaminación que presentaron estos suelos conforme lo establecido por la EPA (1996) en el método 4030, que considera como alto

aquellas concentraciones superiores a 1,000 mg kg⁻¹ de TPH en suelo seco, y un nivel medio a las concentraciones entre 100 y 1,000 mg kg⁻¹ de TPH en suelo seco, y baja a las concentraciones que sean menores a 100 mg kg⁻¹ de TPH en suelo seco.

6. RESULTADOS Y DISCUSIÓN

En la primera parte se analizan los resultados que se obtuvieron de la caracterización físico-química de las muestras; en la segunda parte se presentan los resultados obtenidos del modelo de clasificación para identificar suelos agrícolas contaminados con compuestos derivados del petróleo; en la tercera parte se muestran los resultados generados por el modelo de predicción para cuantificar la contaminación de los suelos agrícolas con petróleo crudo y finalmente en la última parte se realiza el estudio de la contaminación de los suelos de la Cuenca Baja del Río Tonalá.

6.1. Caracterización Físico-Química de los Suelos

La caracterización físico-química de los suelos (Anexo, Cuadro 3) indica un pH ácido en la mayoría de ellos. Dependiendo del sitio de muestreo los niveles del pH disminuyen, presentando en la mayoría de los suelos el pH más alto en los sitios con derrame descendiendo hacia la desembocadura de la cuenca. A excepción de los Arenosoles que presentan un pH neutro, y los cambios de pH presentan contenidos más bajo en los sitios con derrame, y los contenidos más altos en la desembocadura.

Para el caso de la conductividad eléctrica se encontró que los Gleysoles en el sitio del derrame presentan una salinidad muy fuerte, y a partir del sitio de orilla de campo hacia la desembocadura y el testigo, presenta una conductividad eléctrica con un efecto despreciable de la salinidad. Los suelos Histosoles y el Solonchak son fuertemente salinos; a diferencia de estos, los Arenosoles presentan una salinidad baja.

Estos suelos presentan contenidos muy altos de materia orgánica a excepción de los Arenosoles, en los cuales los contenidos son de bajos a medios; el nivel de nitrógeno total es de bajo a medio en los Arenosoles y las muestras de Gleysoles, a muy altos en las muestras de Histosoles y Solonchak. En el caso de las bases intercambiables se encontró que según la matriz de suelo se presentan concentraciones diferentes: en el suelo Histosol el sodio es el más alto y los más bajos corresponden al catión potasio, a diferencia de la matriz de suelos Histosol fibrico el cual presenta una concentración alta de calcio y los más bajos corresponden al potasio; el suelo Solonchak presenta contenidos altos de sodio y los más bajos los presenta el potasio; los Gleysoles y Arenosoles presentan concentraciones altas de calcio y las más bajas corresponden al potasio. La capacidad de intercambio catiónico que presentan estos suelos es muy baja en los Arenosoles, de medio a alto en el Gleysol y muy alta en los Histosoles y Solonchaks. La textura de los suelos son de las clases: arenas en los arenosoles, y migajones arcillo arenoso y arcillas en Gleysoles y solonchaks.

Los Histosoles son suelos orgánicos, ricos en nutrimentos, con baja densidad aparente, alta retención de humedad, manto freático elevado, presentan salinidad y sodicidad alta; según la clasificación de uso se tierras se ubican en la séptima clase de uso, es decir, sólo sirven para la actividad forestal y como reserva de vida silvestre; se clasificaron las subunidades H. fibrico (HSf), H. térrico (HSs) (Palma *et al.*, 2007 modificado en el 2009).

Los Gleysoles presentan inundaciones frecuentes y manto freático elevado; el pH es ácido y presentan salinidad y sodicidad. Se localizaron las subunidades G. verti-eútrico (GLev), G. mólico (GLm) y G. úmbrico (GLu).

El Solonchak gléyico (SCg) es un suelo de séptima clase, salino y sódico, con inundaciones frecuentes y manto freático elevado; sólo apto para la actividad

forestal y de protección de vida silvestre, según la clasificación de uso de tierras.

Los Arenosoles tienen textura arenosa, encharcamiento ocasional (A. gléyico), baja fertilidad y manto freático elevado.

6.2. Interpretación de los Espectros NIRS de los Suelos

Con las lecturas de las 76 muestras de suelo analizadas con el NIRS se obtuvo un dato por cada dos longitudes de onda, y estas, se presentan en números pares desde la longitud de onda 400 hasta los 2500 nm; con lo cual, se obtuvo un total de 1,050 registros, datos o variables por muestra. Por tanto, se obtuvo una matriz de datos conformada por un total de 79,800 datos, los cuales fueron representados en los espectros que se observan en la Figura 15.

La Figura 15 contiene el intervalo de longitudes de onda (eje de las abcisas) y el logaritmo del inverso de los registros (eje de las ordenadas). La presentación del logaritmo el inverso de los registros (\log_1/R) está definida en el programa ISISCAN.

En dicha Figura se puede observar que en la región espectral de 400 a aproximadamente 1100 nm, correspondiente a la región del UV-Vis, los espectros de las muestras presentan muchas vibraciones y por ende no se le puede dar un seguimiento a cada una de las muestras. Sin embargo, a partir de la longitud de onda 1100, los espectros presentan un mejor comportamiento y se observa mayor unión entre muestras, a excepción de dos de ellas (30 y 33) que se separan del grupo

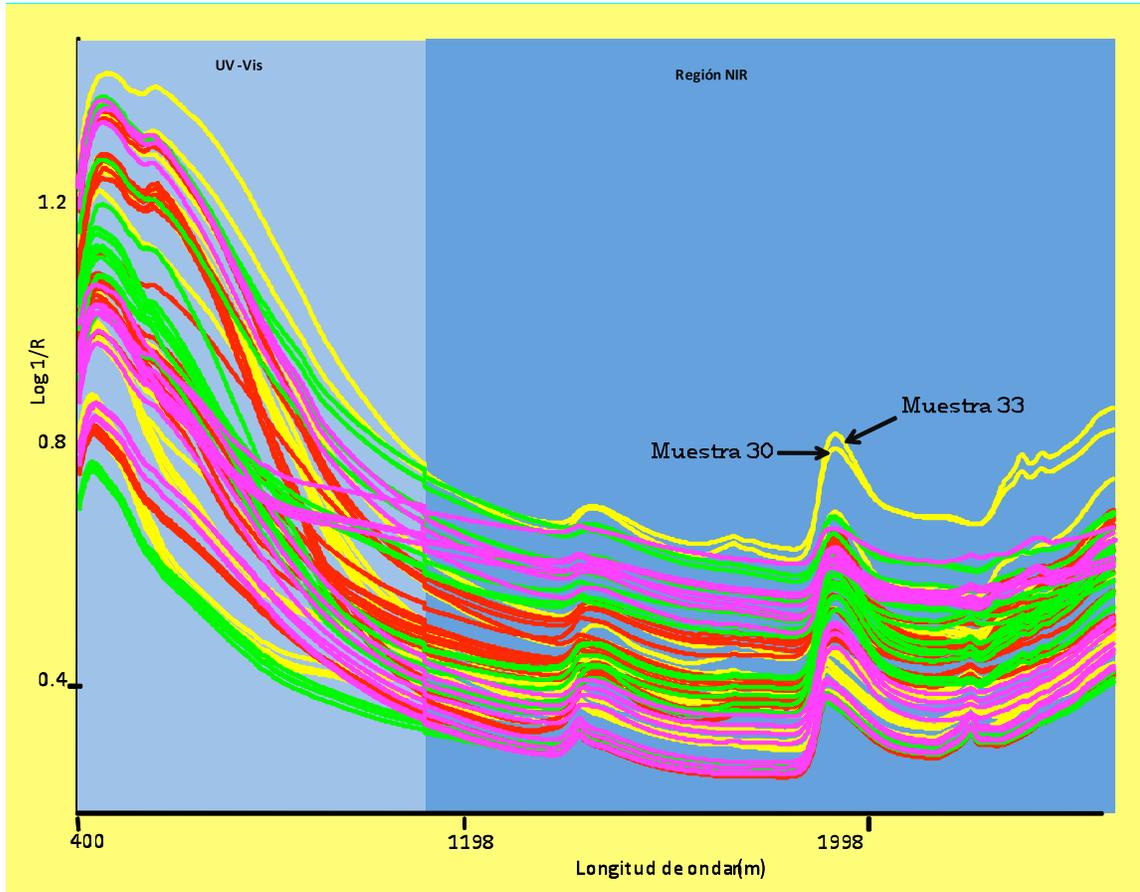


Figura 15. Espectro total Nirs de 76 muestras de suelos procedentes de la Cuenca Baja del río Tonalá.

Con la selección de la parte del infrarrojo cercano, del espectro original de las muestras (Figura 16) se puede observar que en esta región existe uniformidad en el espectro del conjunto de las muestras, también se observan picos bien definidos en las diferentes longitudes de onda lo cual es debido al estrechamiento y ensanchamiento de bandas. Se identificó un primer pico (1) a la longitud de onda 1440 nm que pertenece al grupo de los metilenos (CH-CB-CH₂). En la longitud de onda 1726 nm se encontró un segundo pico (2) en el cual se pueden encontrar grupos de compuestos de metilenos con enlaces carbono-oxígeno y por ensanchamiento de las bandas del primer enlace de grupos metileno (C-O stretch 1st CH₂); en la misma

figura se observó que aproximadamente en la longitud de onda 1928 a 1932 nm se presenta un tercer pico (3) de grupos con enlaces OH, los cuales pueden pertenecer a grupos formadores del almidón y la celulosa.

El cuarto pico que se observa en el intervalo de las longitudes de onda 1196-2206 nm, en algunas de las muestras, pertenece a compuestos con ensanchamiento de bandas en los enlaces C-H (C-H stretch /C=O CB--CHO) y de compuestos con doble enlace carbono-oxígeno C=O (C-H ensanchamiento /C=O CB—CHO). También se encontraron otros dos picos: (5) a los 2306 y 2308 nm los cuales pertenecen a compuestos de grupos de las proteínas y algunos aceites (C-H bend 2nd OT-protein, CH bend 2nd OT-oil), y el sexto pico (6) que se encuentra a partir de los 2484 nm se presenta debido a un ensanchamiento en las bandas de los enlaces C-H (C-H ensanchamiento/C-C CB-celulosa). También se encuentran grupos con enlaces C-C pertenecientes a los formadores de la celulosa.

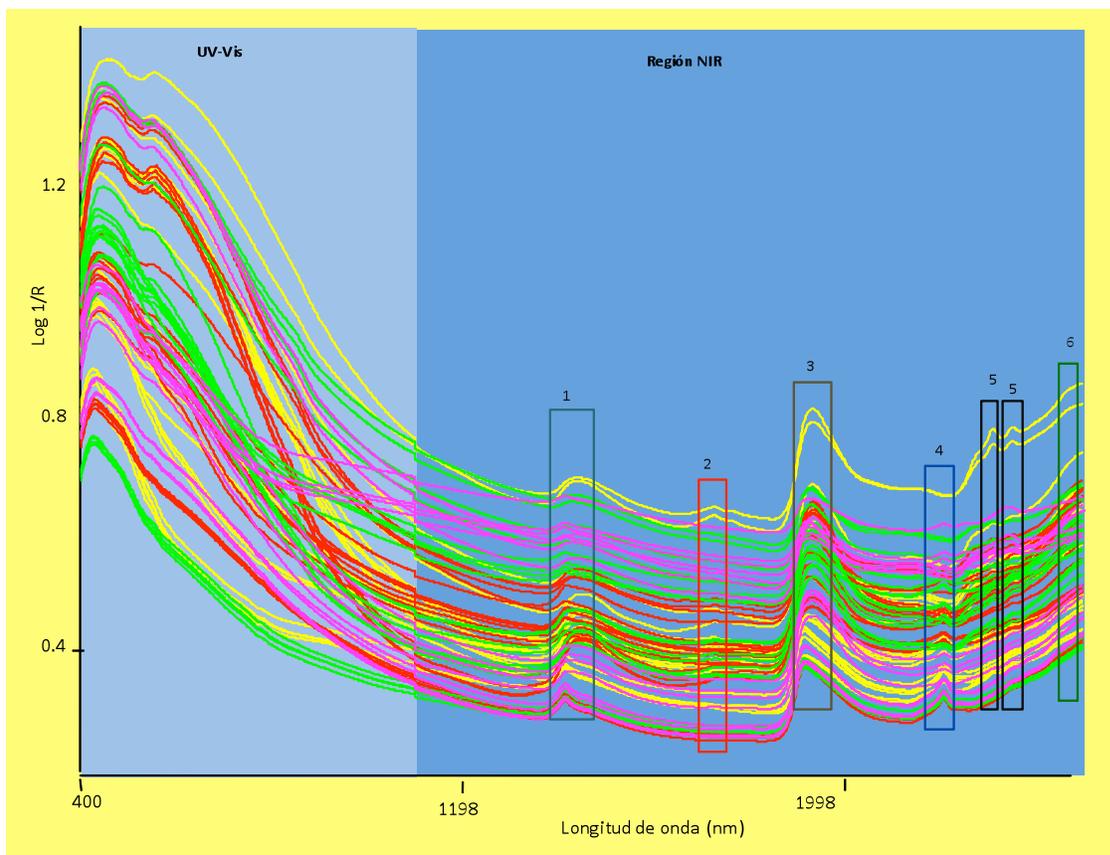


Figura 16. Selección región NIR

6.3. Análisis de Componentes Principales (PCA)

El primer análisis quimiométrico de la señal espectral de las 76 muestras, analizadas con la tecnología NIRS, fue realizada con técnicas no supervisadas, como es análisis de componentes principales (PC), las cuales se aplicaron con objeto de identificar las estructuras latentes que no es posible observar en una lectura directa de la matriz de datos (76X1050). El objetivo de esta técnica consiste en identificar el menor número de componentes principales que explique la mayor parte de la varianza de las observaciones de modo que los residuales sean los más pequeños posibles.

En el Cuadro 5, se presentan los resultados de las tres primeras componentes principales, los cuales suman el 98.27% de la varianza explicada. Se incluyen los valores propios, la varianza explicada, el porcentaje de varianza explicada y el porcentaje de varianza acumulada, de PC1, PC2 y PC3.

Cuadro 5. Valores representativos de las tres primeras componentes principales del espectro UVis-NIR.

	PC 1	PC 2	PC 3
Valor propio	185.06	101.52	28.87
Varianza explicada	189.65	96.94	28.87
Porcentaje de varianza explicada	59.08	30.20	8.99
Porcentaje acumulado de varianza explicada	50.08	89.28	98.27
Porcentaje de varianza residual	40.92	10.72	1.73

PC= Componente principal.

La varianza total explicada por las componentes principales 1, 2 y 3; corresponde a 630, 305 y 76 variables originales, respectivamente. Lo cual es debido a lo correlacionadas que están cada una de estas variables con el componente principal respectivo. En otras palabras las 104 variables originales pueden ser representadas únicamente por tres nuevas variables o componentes principales. El porcentaje de la varianza residual, o varianza no explicada, por cada uno de las tres componentes principales es 40.92%, 10.72% y 1.73%, respectivamente. La varianza residual total corresponde a 39 variables, registros o datos generados con el NIRS y que no fueron explicados por las tres componentes principales; variables que son explicadas en las 1,047 componentes principales restantes.

La representación de las 76 muestras de suelo, en el plano definido por las dos primeras componentes principales y que explican el 89.28 % de la varianza acumulada, se muestran en la Figura 19. En la figura se observa que las muestras se distribuyen a lo largo de los PC1 y PC2, sin presentar el

arreglo espacial esperado (ordenamiento de las muestras en función de su origen, tipo de suelo y de la presencia de hidrocarburos derivados del petróleo).

Los valores de las componentes principales (PC1 y PC2) se incluyen en el Cuadro 5, en el cual podemos confirmar que existen valores con amplios intervalos, lo que confirma lo observado en la Figura 17 con la distribución de las muestras a lo largo de las dos componentes principales.



Figura 17. Representación de las puntuaciones de las dos primeras componentes principales del espectro total NIR de las muestras de suelo.

Adicional a lo anterior se realizaron otras combinaciones de las componentes principales (PC1-PC3, PC2-PC3, etc.) sin embargo los resultados fueron similares a los observados en la Figura 17.

Como alternativa a lo anterior, y con objeto de definir una estrategia para la interpretación de las señales espectrales del NIRS se decidió eliminar la región del espectro que corresponde a las longitudes de onda UV-Vis (400-1,100), y utilizar en el siguiente análisis quimiométrico, la región correspondiente al infrarrojo cercano (NIRS). La eliminación de esta región (UV-Vis) fue debido a que esta región presenta mucha vibración y no se le puede dar seguimiento al espectro de cada muestras. Sin embargo con únicamente tres componentes principales se explica casi el 96% de la varianza total.

6.3.1. Análisis de componentes principales de la región espectral NIR

Una vez seleccionada la región del espectro NIR, quedan 700 registros o datos por muestra, con una matriz de 76x700 comprende 53,200 datos. Esto significa una reducción del 33.33 % en el total de datos o registros iniciales.

Las señales espectrales correspondientes recibieron un tratamiento previo al análisis de componentes principales, que consistió en un pre-procesado mediante la media central y la transformación de normalización del área o también conocida como la división de los datos entre la variable 1.

En el Cuadro 6 se muestran los valores representativos de las componentes principales para la región del espectro NIR. La variabilidad acumulada de las tres primeras componentes principales fue del 95.94 %. La varianza total explicada por las componentes principales 1, 2 y 3 corresponde a 418, 203 y 51 variables originales, respectivamente. Lo cual es debido a lo

correlacionada que está cada una de estas variables con la componente principal respectiva. La varianza residual, o varianza no explicada, por cada uno de las tres componentes principales es 40.13, 11.30 y 4.06, respectivamente.

Cuadro 6. Valores representativos de las tres primeras componentes principales de los datos del espectro NIR.

	PC 1	PC 2	PC 3
Valor propio	0.00007	0.000069	0.000011
Varianza explicada	0.000094	0.000045	0.000011
Porcentaje de varianza explicada	59.69	29.00	7.24
Porcentaje acumulado de varianza explicada	59.69	88.70	95.94
Porcentaje de varianza residual	40.13	11.30	4.06

PC= Componente principal.

La varianza residual total o varianza no explicada total corresponde a 28 variables originales generadas con el NIRS; variables que son explicadas en las 697 componentes principales restantes.

La representación de las 76 muestras en el plano definido por las dos primeras componentes principales y que explican el 85.69 % de la varianza acumulada, se muestran en la Figura 18. En la figura se observa que las muestras continúan sin presentar el arreglo espacial esperado (ordenamiento de las muestras en función de su origen y tipo de suelo).

No obstante la selección del espectro NIRS, al realizar el análisis de los componentes principales no fue posible obtener los agrupamientos esperados de las muestras (Figura 18); aún cuando los tres primeros componentes principales son explicativos del fenómeno estudiado, por la variación explicada. Con lo cual se descarta que la señal correspondiente de la región espectral UV-Vis, sea el único factor que interfiere en la formación de los agrupamientos esperados

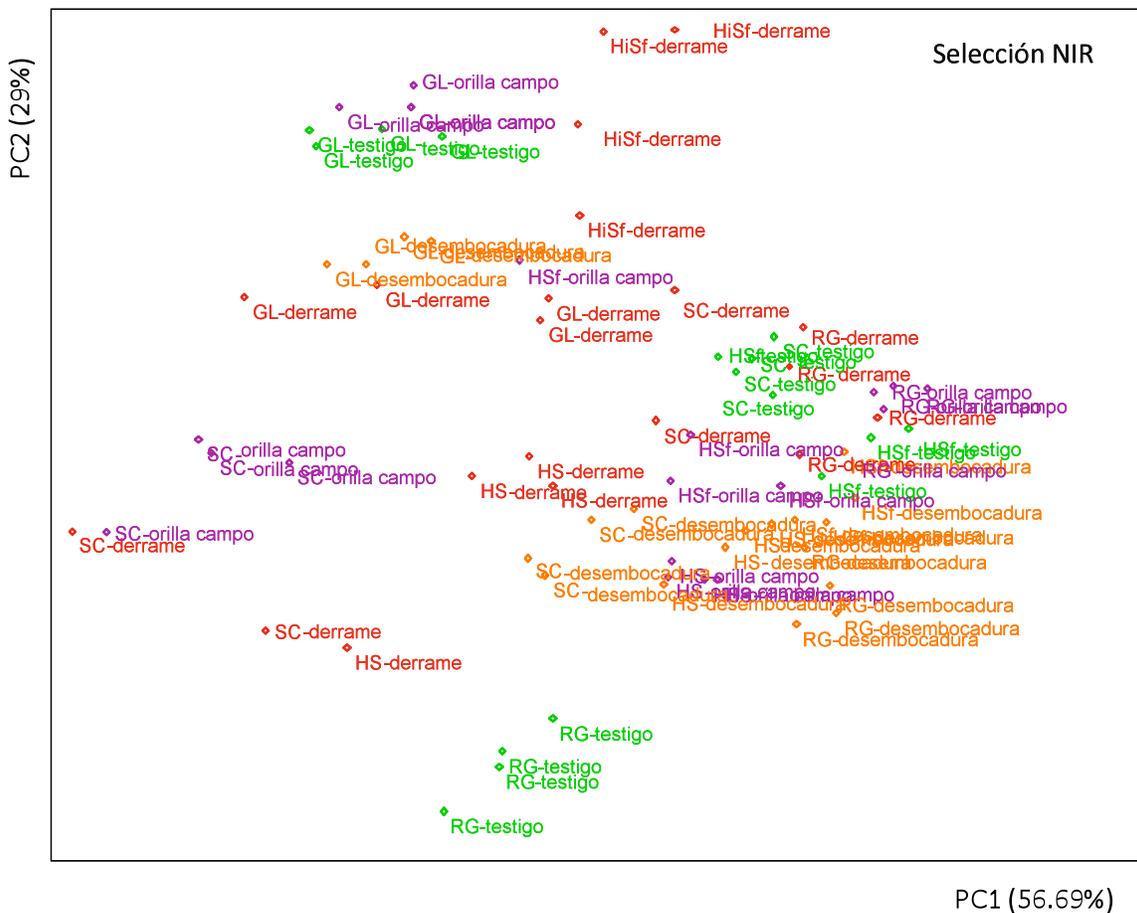


Figura 18. Representación de las puntuaciones de las dos primeras componentes principales del espectro NIR, de las muestras de suelo

Debido a que con el tratamiento quimiométrico anterior no se lograron agrupar las muestras por el tipo de matriz o por el origen de muestreo, se optó por separar las muestras por tipo de suelo y realizar el análisis de las componentes principales de los datos en la región espectral NIR, para ver si se logra encontrar las agrupaciones deseadas.

3.2. Histosol

Los espectros NIR de las 28 muestras de suelo clasificado como Histosol se interpretaron mediante componentes principales. Este análisis se realizó en las 700 variables, lo cual corresponde a una matriz de 28x700 con un total de 19,600 datos. En el Cuadro 7 se presentan los resultados de las tres primeras componentes principales; el tratamiento matemático que se realizó fue un preprocesado de los datos con una media central y una transformación de la división entre el valor de la variable 1. El porcentaje de varianza explicada fue del 99.6 %, en los tres primeros componentes principales.

Cuadro 7. Valores representativos de los tres primeros componentes principales de los datos del espectro NIR de las muestras del suelo Histosol

	PC 1	PC 2	PC 3
Valor propio	35.97	10.86	0.37
Varianza explicada	36.01	10.82	0.38
Porcentaje de varianza explicada	76.04	22.85	0.80
Porcentaje acumulado de varianza explicada	76.04	98.89	99.69
Porcentaje de varianza residual	23.96	1.11	0.31

PC= Componente principal.

La varianza total explicada por los componentes principales 1 y 2 corresponde a 532 y 160 variables originales, respectivamente. Lo cual es debido a lo correlacionada que está cada una de estas variables con el componente principal respectivo. La varianza residual acumulada, o varianza no explicada acumulada, por los dos primeros componentes principales es 23.96 y 1.11, respectivamente.

La representación de las 28 muestras en el plano definido por las dos primeras componentes principales y que explican el 98.89 % de la varianza acumulada, se muestran en la Figura 19. En la Figura se observa que las

muestras presentan un arreglo espacial esperado en función del tipo de suelo y su origen (sitio de muestreo).



Figura 19. Representación de las puntuaciones de las dos primeras componentes principales del espectro correspondiente a la región NIR de las muestras clasificadas como Histosol.

Las puntuaciones de las dos primeras componentes principales (Figura 19) dividen las muestras en dos grandes grupos, grupo de Histosol fibrico (circulo punteado color celeste) y grupo de Histosol Sodic (circulo punteado color rosa); siendo el PC1 el que más contribuye a la separación de estos dos grupos. Dentro de los dos grupos es posible observar la agrupación de las muestras en función de su origen (sitio de muestreo); siendo el PC2 el que más contribuye a la formación de estos subgrupos. Estos subgrupos

corresponden a los sitios de muestreo, testigo, desembocadura, orilla del campo y sitio del derrame.

En una interpretación general de la Figura 19, se observa que las puntuaciones del PC2 (Anexo, Cuadro 4) tienen valores, de mayor a menor, en el orden siguiente: muestras de la desembocadura, muestras testigo, muestras de orilla del campo y muestras del sitio de derrame de petróleo crudo, estando perfectamente diferenciados cada subgrupo a excepción de una muestra del sitio testigo que se encuentra cerca de las muestras de la orilla del campo. La decisión de comentar los valores de las puntuaciones del PC2 es por la contribución que tiene en la formación de los subgrupos.

La separación tan marcada de las muestras del sitio de derrame, puede atribuirse a la presencia de contaminantes derivados del petróleo en una concentración mayor a la que se encuentra en los otros sitios de muestreo.

6.3.3. Gleysol

Los espectros NIR de las 16 muestras de suelo clasificado como Gleysol se interpretaron mediante componentes principales. Este análisis se realizó en las 700 variables, lo cual corresponde una matriz de 16x700 con un total de 11,200 datos. En el Cuadro 8 se presentan los resultados de las tres primeras componentes principales; para el tratamiento matemático se realizó un preprocesado de los datos con una media central y una transformación de la división entre el valor de la variable 1. El porcentaje de varianza explicada fue del 99.91 %, en las tres primeras componentes principales (95.49, 3.81 y 0.61, respectivamente).

Cuadro 8. Valores representativos de las tres primeras componentes principales de los datos del espectro NIR de las muestras de suelo Gleysol.

	PC 1	PC 2	PC 3
Valor propio	3.52	2.04	0.03
Varianza explicada	5.36	0.21	0.03
Porcentaje de varianza explicada	95.49	3.81	0.61
Porcentaje acumulado de varianza explicada	95.49	99.30	99.91
Porcentaje de varianza residual	4.51	0.70	0.09

PC= Componente principal.

La varianza total explicada por las componentes principales 1 y 2 corresponde a 668 y 27 variables originales, respectivamente. Por la correlación que tiene cada una de estas variables con el componente principal respectivo. La varianza residual acumulada, o varianza no explicada acumulada, por los dos primeros componentes principales es 4.51 y 0.70, respectivamente.

La representación de las 16 muestras en el plano definido por las dos primeros componentes principales y que explican el 98.89 % de la varianza acumulada, se muestran en la Figura 20. En la figura se observa que las muestras presentan el arreglo espacial esperado en función de su origen (sitio de muestreo).

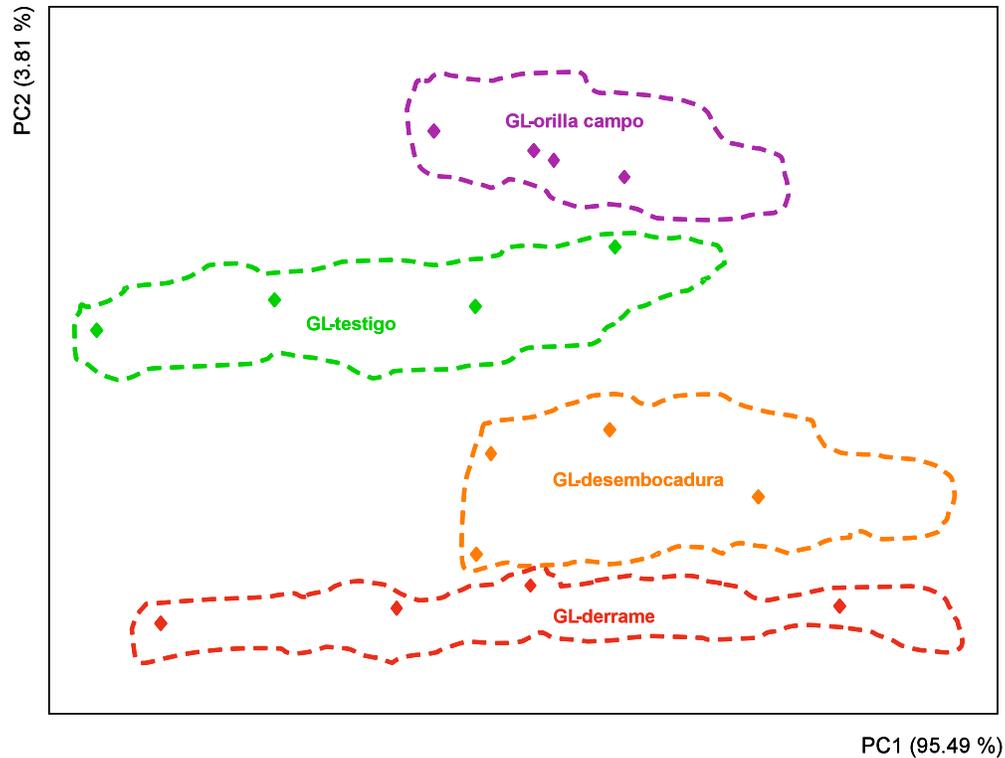


Figura 20. Representación de las puntuaciones de las dos primeras componentes principales del espectro correspondiente a la región NIR de las muestras clasificadas como Gleysol.

Las puntuaciones de las dos primeras componentes principales (Figura 20) dividen las muestras en cuatro grupos; con una distribución de las muestras a lo largo de la PC1. Se observa que es la PC2 quien más contribuye a la separación de los cuatro grupos.

En la Figura 20 se observa que las puntuaciones de la PC2 tienen valores, de mayor a menor, en el orden siguiente: muestras de orilla del campo, muestras testigo, muestras de la desembocadura y muestras del sitio de derrame de petróleo crudo. Las muestras colectadas en la parte de la desembocadura de la cuenca del río, se encuentran cercanas a las del sitio del derrame, esto puede explicarse porque, en el sitio del derrame se realizó un tratamiento de limpieza del suelo y se adicionó suelo procedente de otro sitio.

Las muestras testigo y las de la orilla del campo se separan perfectamente de las muestras de la desembocadura y del sitio del derrame; lo cual, puede atribuirse a la presencia de contaminantes derivados del petróleo en una mayor concentración en las muestras procedentes de los sitios del derrame y desembocadura, a la que se encuentra en los dos primeros sitios de muestreo.

6.3.4. Solonchack

Los espectros NIRS de las 16 muestras de suelo clasificado como Solonchack se interpretaron mediante componentes principales. Este análisis se realizó en las 700 variables, lo cual corresponde a una matriz de 16x700 con un total de 11,200 datos. En el Cuadro 9 se presentan los resultados de las tres primeras componentes principales; para su análisis matemático se realizó un preprocesado de los datos con una media central. El porcentaje de varianza explicada fue del 99.93 %, en las tres primeras componentes principales.

Cuadro 9. Valores representativos de las tres primeras componentes principales de los datos del espectro NIRS de las muestras de suelo Solonchack.

	PC 1	PC 2	PC 3
Valor propio	54.21	46.96	0.14
Varianza explicada	99.93	1.24	0.14
Porcentaje de varianza explicada	98.57	1.22	0.14
Porcentaje acumulado de varianza explicada	98.57	99.80	99.93
Porcentaje de varianza residual	1.43	0.20	0.07

PC= Componente principal.

La varianza total explicada por las componentes principales 1 y 2 corresponde a 690 y 8 variables originales, respectivamente. La varianza residual acumulada, o varianza no explicada acumulada por las dos primeras componentes principales es 1.43 y 0.20, respectivamente.

La representación de las 16 muestras en el plano definido por las dos primeras componentes principales y que explican el 99.79 % de la varianza acumulada se muestra en la Figura 21. En la figura se observa que las muestras presentan un arreglo espacial esperado en función de su origen (sitio de muestreo).

Los datos obtenidos en el espectro para este grupo de suelo, tuvieron un pre-procesado en la media central, presenta en sus tres primeras componentes principales una varianza acumulada del 99.93 % (98.57, 1.222 y 0.13, respectivamente).

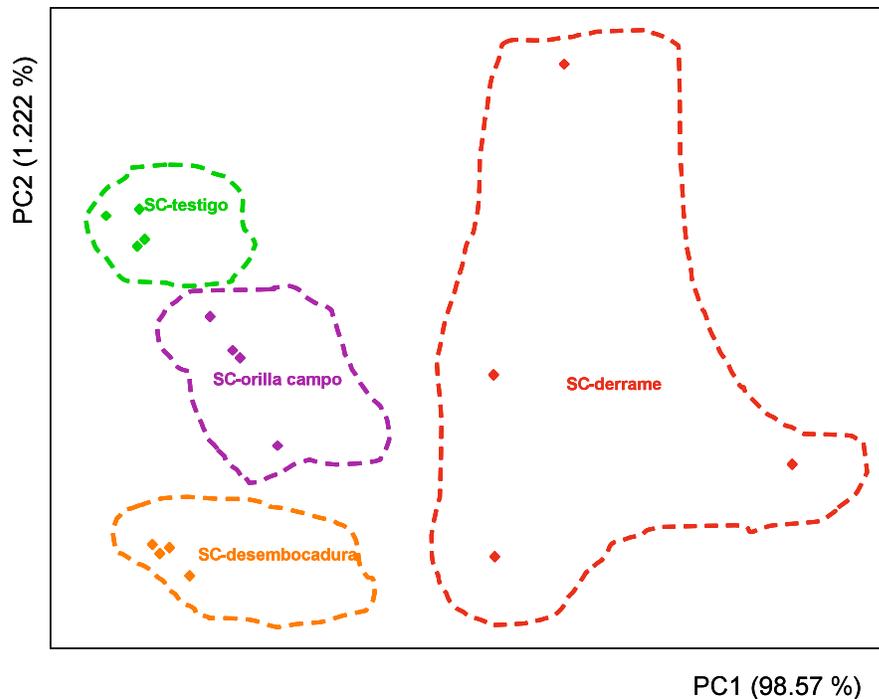


Figura 21. Representación de las puntuaciones de las dos primeras componentes principales del espectro correspondiente a la región NIRS de las muestras clasificadas como Solonchack.

Las puntuaciones de las dos primeras componentes principales (Figura 21) dividen las muestras en cuatro grupos; con participación de las dos componentes principales en la distribución y agrupamiento de las muestras. En los grupos formados es posible observar la agrupación de las muestras en función de su origen (sitio de muestreo); estos grupos corresponden a los sitios de muestreo testigo, orilla del campo, desembocadura y sitio del derrame.

En la Figura 21 se observa que las puntuaciones de la PC2 tienen valores, de mayor a menor, en el orden siguiente: muestras testigo, muestras de orilla del campo y muestras de la desembocadura, mientras que las muestras del sitio de derrame de petróleo crudo se distribuyen a lo largo de la PC2; por lo que esta componente solo contribuye a la formación de agrupamientos para las muestras testigo, orilla del campo y desembocadura. La separación del agrupamiento de las muestras del sitio del derrame de los tres agrupamientos restantes es contribución de la CP1.

La separación de los agrupamientos de las muestras testigo, orilla del campo y desembocadura, y el agrupamiento de muestras del sitio del derrame, se puede atribuir a la presencia de contaminantes derivados del petróleo en una mayor concentración a la que se encuentra en los tres primeros sitios de muestreo.

6.3.5. Arenosol

Los espectros NIR de las 16 muestras de suelo clasificado como Arenosol se interpretaron mediante componentes principales. Este análisis se realizó en las 700 variables, lo cual corresponde a una matriz de 16x700 con un total de 11,200 datos. En el Cuadro 10 se presentan los resultados de las tres primeras componentes principales; para este análisis se realizó un

preprocesado de los datos con una media central y una transformación de las variables divididas en la variable 1. El porcentaje de la varianza explicada es del 99.96 %, en las tres primeras componentes principales (99.62, 0.27 y 0.074, respectivamente)..

La varianza total explicada por las componentes principales 1 y 2 corresponde a 697 y 2 variables originales, respectivamente. La varianza residual acumulada, o varianza no explicada acumulada, por las dos primeras componentes principales es 0.38 y 0.11, respectivamente.

Cuadro 10. Valores representativos de las tres primeras componentes principales de los datos del espectro NIRS de las muestras de suelo Arenosol.

	PC 1	PC 2	PC 3
Valor propio	29.48	26.21	0.04
Varianza explicada	55.55	0.15	0.04
Porcentaje de varianza explicada	99.62	0.27	0.074
Porcentaje acumulado de varianza explicada	99.62	99.89	99.96
Porcentaje de varianza residual	0.38	0.11	0.04

PC= Componente principal.

La representación de las 16 muestras en el plano definido por las dos primeras componentes principales y que explican el 99.89 % de la varianza acumulada, se muestran en la Figura 22. Y se observa que las muestras presentan el arreglo espacial esperado en función de su origen (sitio de muestreo).

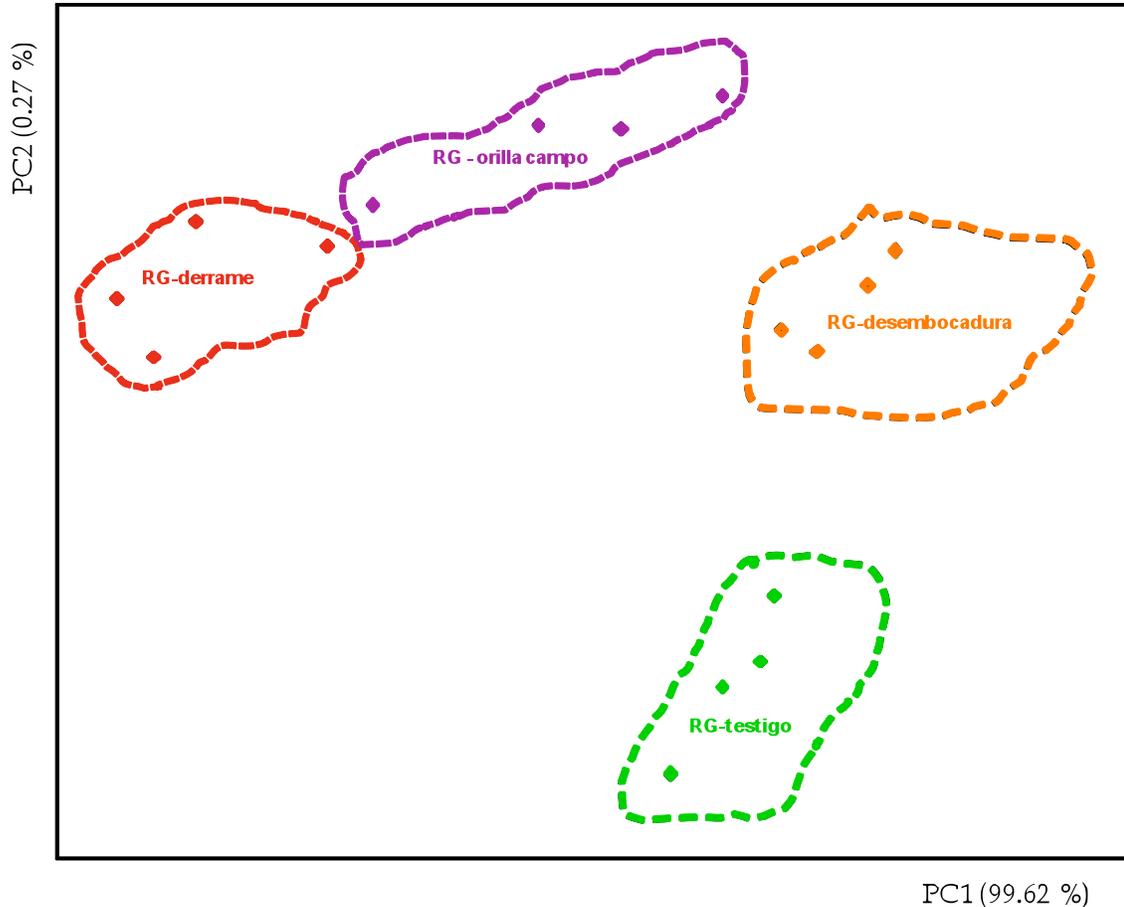


Figura 22. Representación de las puntuaciones de las dos primeras componentes principales del espectro correspondiente a la región NIR de las muestras clasificadas como Arenosol.

Las puntuaciones de las dos primeras componentes principales (Figura 22) dividen las muestras en cuatro grupos; los agrupamientos están en función del origen de las muestras (sitio de muestreo). Estos grupos corresponden a los sitios de muestreo testigo, orilla del campo, desembocadura y sitio del derrame. Las dos componentes principales contribuyen en la formación de los agrupamientos de las muestras.

La separación de los agrupamientos de las muestras se puede atribuirse a la presencia de contaminantes derivados del petróleo, en la muestra del sitio de

derrame tiene una concentración mayor a la que se encuentra en los tres sitios restantes.

6.4. Modelado Independiente de Clases (SIMCA)

Debido al efecto matricial observado durante el análisis exploratorio con las componentes principales, se realizó el análisis del comportamiento de las muestras en el modelo de clasificación esto para cada uno de los suelos.

En la matriz de datos espectrales, generados con el NIRS, primero se realizó un preprocesado con lo cual se pudo observar que las muestras se separan adecuadamente, pauta que indica la alta probabilidad de éxito en la realización de un modelo de clasificación.

6.4.1. Modelos de clasificación SIMCA

Se generaron modelos para cada uno de los tipos de suelo al igual que en el análisis de componentes principales, a los datos originales (generados con el NIRS) se les realizó diferente preprocesado y transformación; la frontera de cada clase se estableció por medio de una distancia crítica con un nivel de significancia del 5%.

6.4.1.1. Modelo de clasificación SIMCA para el Solonchak

Para la discusión del modelo clasificación SIMCA se tomaron como referencia los gráficos de Coomans generados con los espectros NIRS del Solonchack (Figura 23). El modelado de la matriz Solonchak se realizó mediante un preprocesado de los datos de la varianza escalar y una

transformación aplicando la segunda derivada de los datos. La distancia para establecer las fronteras entre las clases con un nivel de significancia del 5%. En la revisión de literatura fue descrito el gráfico de Coomans, en cuanto a la distribución de sus clases, y será este formato el utilizado para la discusión de resultados.

En la Figura (23a) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio orilla del campo, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases testigo y desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

En la Figura (23b) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio desembocadura, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases testigo y orilla del campo, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

En la Figura (23c) se observa que al comparar la clase B, correspondiente a las muestras de suelo del sitio desembocadura, con la clase A, correspondiente a las muestras de suelo del sitio orilla del campo; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases testigo y derrame, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante

superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

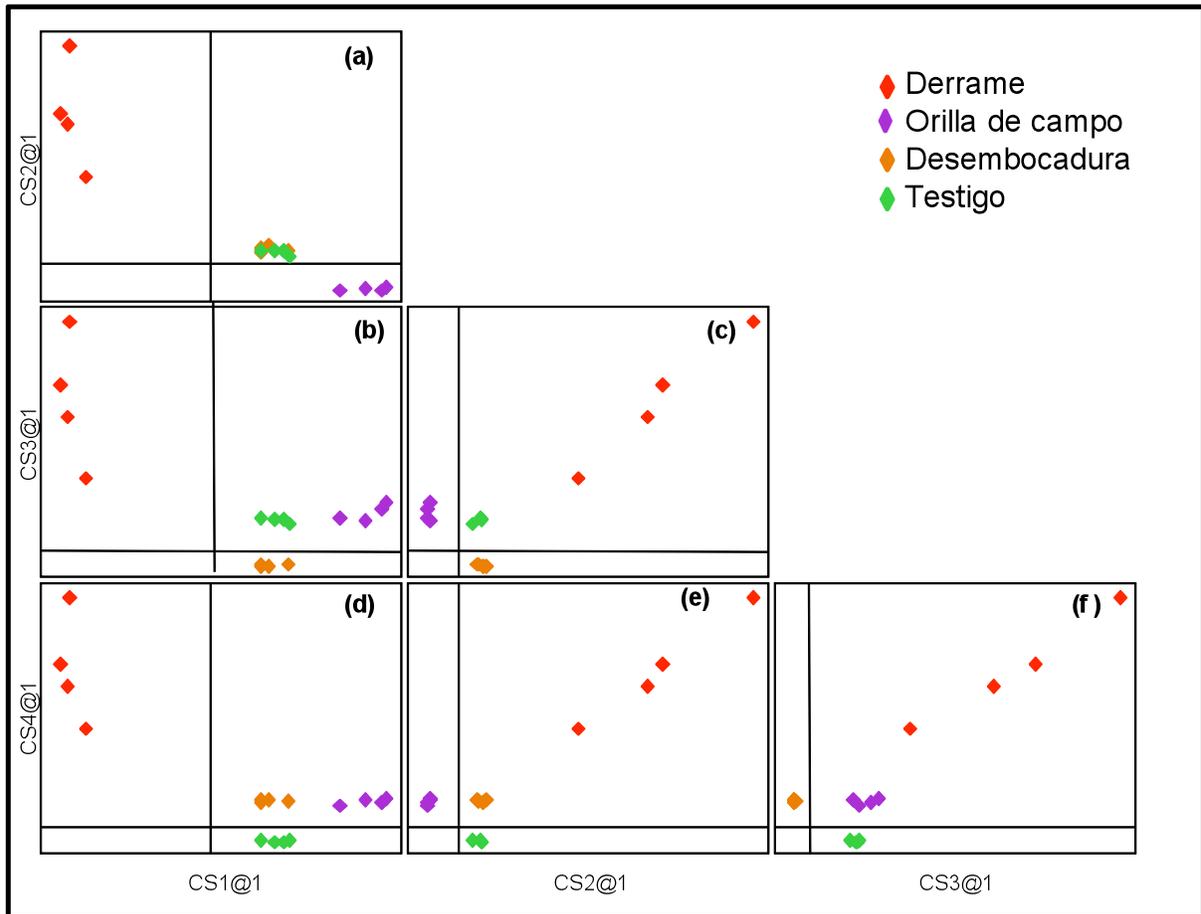


Figura 23. Diagrama de Coomans construido con el modelado de clases SIMCA, para las muestras de suelo Solonchack.

En la Figura (23d) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases orilla de campo y desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el

cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

En la Figura (23e) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio orilla del campo; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases derrame y desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

Por último, en la Figura (23f) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio desembocadura; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases orilla del campo y derrame, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

En la Figura 23, se observan las proyecciones de clases, del suelo Solonchack se observó que los sitios de muestreo testigo y desembocadura se localizan en las mismas coordenadas de la figura. Así mismo, en la Figura 23, en el cuadrante superior derecho del gráfico de Coomans (a) (correspondiente a la zona de muestras desconocidas); se ubican las muestras de los sitios desembocadura y testigo, con coordenadas similares, lo cual indica que existe una semejanza en la composición global de las muestras. Sin embargo, en el gráfico de Coomans (f), en el cual se comparan las muestras de los sitios desembocadura y testigo, estos grupos de muestras forman sus clases respectivas; esto significa que aún con la

posible similitud manifestada en las figuras previas, las muestras de la clase B son las del sitio testigo y las muestras de la clase A son las del sitio desembocadura.

Con los resultados observados en la Figura 23, al realizar el comparativo entre clases, ninguna de ellas se mezcla con otra, ni se salen del área de acoplamiento, es decir, que la sensibilidad de los modelos son del 100%; dado que las cinco clases de muestras son incluidas en las clases respectivas (A ó B) cuando son utilizadas para construir los modelos.

Lo discutido en el párrafo anterior sirve para precisar que aún cuando las muestras del Solonchack, de los sitios de muestreo desembocadura y testigo, presentan similitud en sus propiedades físicas y químicas, mediante el uso de la tecnología NIRS es puede generar información útil y suficiente que nos permita establecer modelos de clasificación mediante técnicas quimiométricas (SIMCA).

6.4.1.2. Modelo de clasificación SIMCA para el Arenosol

El modelado de la matriz de suelo Arenosol (Figura 24) se realizó mediante un preprocesado de los datos en el rango de escala y una transformación de los mismos aplicando la segunda derivada y la división de la muestra del vector cinco.

En la Figura (24a) se puede ver que al comparar la clase B, correspondiente a las muestras de suelo del sitio orilla del campo, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases testigo y desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B), son colocadas en el

cuadrante superior derecho, que corresponde a muestras desconocidas. Las mismas muestras están claramente separadas una de la otra y se puede observar que no existen muestras con similitud a ambas clases.

En la Figura (24**b**) se observa que al comparar la clase B, correspondiente a las muestras de suelo del sitio desembocadura, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases testigo y orilla del campo, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

En la Figura (**c**) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio desembocadura, con la clase A, correspondiente a las muestras de suelo del sitio orilla del campo; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases testigo y derrame, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

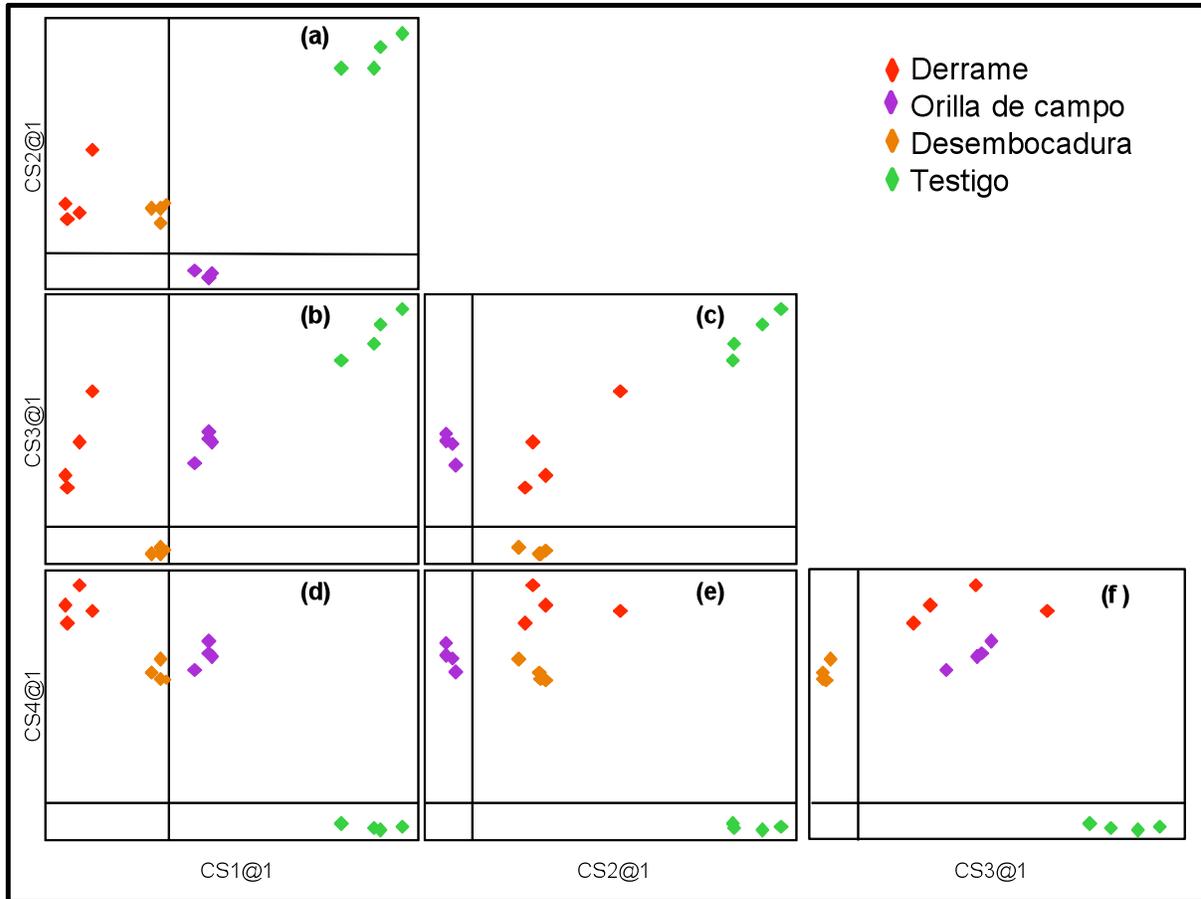


Figura 24. Diagrama de Coomans construido con el modelado de clases SIMCA para las muestras de suelo Arenosol.

En la Figura (24d) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases orilla de campo y desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

En la Figura (24e) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio derrame;

correspondiente a las muestras de suelo del sitio orilla del campo; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases derrame y desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases.

Por último, en la Figura (24**f**) se observa que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio desembocadura; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases orilla del campo y derrame, al no presentar semejanza con ninguna de las clases anteriores son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas, por no pertenecer a las clases A ó B. No existen muestras con similitud a ambas clases.

La Figura 24, muestra la proyección de clases del suelo Arenosol, y se observa que en el cuadrante superior derecho del gráfico de Coomans (**a**) (correspondiente a la zona de muestras desconocidas) se ubican las muestras de los sitios desembocadura y testigo. Lo cual indica que no existe una semejanza en la composición global de las muestras. Sin embargo, en el gráfico de Coomans (**f**), en el cual se comparan las muestras de los sitios desembocadura y testigo, estos grupos de muestras forman sus clases respectivas; esto significa que aún con la posible similitud manifestada en las figuras previas, las muestras de la clase B son las del sitio testigo y las muestras de la clase A son las del sitio desembocadura.

Con los resultados de las comparaciones observadas en la Figura 24, al realizar el comparativo entre clases, ninguna de ellas se mezcla con otra, ni se salen del área de acoplamiento, es decir, que la sensibilidad de los

modelos son del 100%; dado que las cuatro muestras son incluidas en las clases respectivas (A ó B) cuando son utilizadas para construir los modelos.

Lo discutido en el párrafo anterior sirve para precisar que aún cuando las muestras del Arenosol, de los sitios de muestreo desembocadura y testigo, presentan similitud en sus propiedades físicas y químicas mediante el uso de la tecnología NIRS se puede generar información útil y suficiente que nos permite establecer modelos de clasificación mediante técnicas quimiométricas (SIMCA).

6.4.1.3. Modelo de clasificación SIMCA para el Gleysol

La generación del modelo para el suelo Gleysol (Figura 25), se realizó mediante un preprocesado de los datos mediante el rango de escala y una transformación aplicando una división entre la muestra del vector uno.

El modelado de esta matriz de suelo lo podemos observar en la figura (a). En la cual al realizar la comparación entre clases A y B (correspondientes a las muestras del sitio derrame y orilla de campo, respectivamente), las muestras están en sus respectivas áreas de acoplamiento, sin embargo a pesar de esto se observa que las muestras pertenecientes a la clase B (muestras del sitio orilla de campo) se encuentran en la frontera de su área de acoplamiento con la zona de las muestras desconocidas, lo que podemos deducir que estas comparten ciertas características con la muestras de la zona desconocida (sitios testigo y desembocadura).

En la Figura (25b) se muestra que al comparar la clase B, correspondiente a las muestras de suelo del sitio desembocadura, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras

de las clases testigo y orilla del campo, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases. Se observa que las muestras localizadas en la zona de muestras desconocidas están separadas en dos grupos de muestras.

En el gráfico (25c) se observa que al comparar la clase B, correspondiente a las muestras de suelo del sitio desembocadura, con la clase A, correspondiente a las muestras de suelo del sitio orilla del campo; cada una de estas se ubica en su respectiva área de acoplamiento, sin embargo las muestras de la clase A se presentan en el límite de la región de confianza de la clase con respecto a la zona de las muestras desconocidas. Y en el caso de las muestras de las clases testigo y derrame, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B), son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. Estas muestras a pesar de que se presentan en dos grupos, sus fronteras se encuentran cercanas. No existen muestras con similitud a ambas clases.

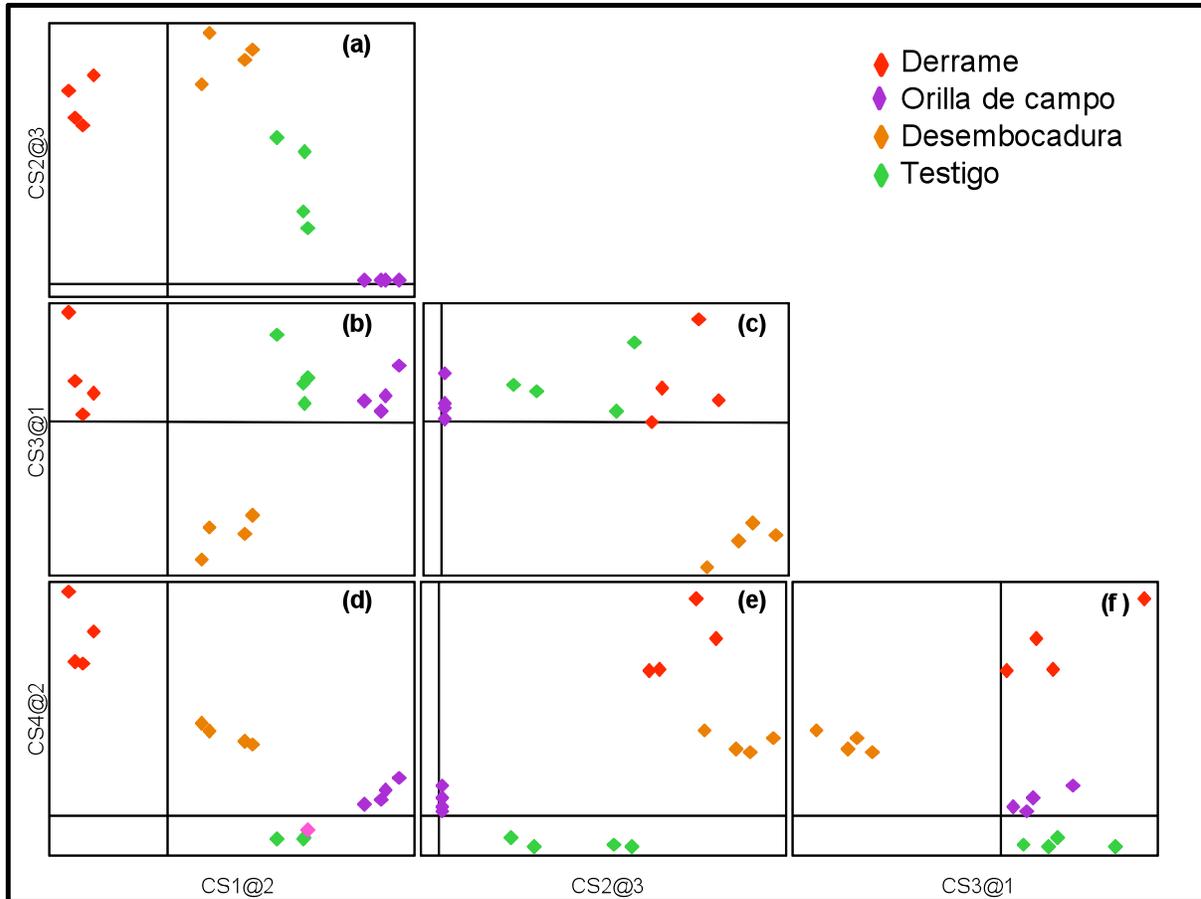


Figura 25. Diagrama de Coomans construido con el modelado de clases SIMCA, para muestras del suelo Gleysol.

Se puede observar que la región de confianza de la clase B es mayor que la de la clase A y de la zona de muestras desconocidas, por lo cual las muestras se ubican más alejadas de la frontera de clase con respecto a la zona de las muestras desconocidas.

En el gráfico (25d) se observa que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases orilla de campo y desembocadura, al no presentar semejanza con ninguna

de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. No existen muestras con similitud a ambas clases. Se puede observar que la distancia de las fronteras entre los grupos de muestras, están separadas entre grupos, tanto en la zona de la clase A y B como en la zona de las desconocidas.

En el área de acoplamiento de las muestras desconocidas, las muestras correspondientes al sitio de orilla de campo están más cercanas al límite de la frontera con la clase B, mostrándose este grupo de muestras más juntas, a diferencia de las muestras del sitio desembocadura que se encuentran en la misma área de acoplamiento pero las muestras están más separadas una de las otras.

En el gráfico (25e) se observa que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio orilla del campo; cada una de éstas se ubica en su respectiva área de acoplamiento, sin embargo las muestras de la clase A se encuentran en el límite de la frontera de la clase con la zona de las muestras desconocidas. Y en el caso de las muestras de las clases derrame y desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas.

Por último, en la Figura (25f) se muestra que al comparar la clase B, correspondiente a las muestras de suelo del sitio testigo, con la clase A, correspondiente a las muestras de suelo del sitio desembocadura; cada una de estas se ubica en su respectiva área de acoplamiento. En el caso de las muestras de las clases orilla del campo y derrame, al no presentar semejanza con ninguna de las clases anteriores son colocadas en el cuadrante superior derecho; que corresponde a muestras desconocidas, con una clara separación entre grupo de muestras. Las muestras de esta zona se

encuentran muy cercanas a la frontera de la clase A y las del sitio orilla de campo también están cercanas la frontera de la clase B. No existen muestras con similitud a ambas clases.

En la Figura 25, se muestra la proyección de clases del suelo Gleysol, y se quebserva, en el cuadrante superior derecho del gráfico de Coomans (**a**) (correspondiente a la zona de muestras desconocidas) se ubican las muestras de los sitios desembocadura y testigo, con diferentes coordenadas. Lo cual indica que no existe una semejanza en la composición global de las muestras. Sin embargo, en el gráfico de Coomans (**f**), en el cual se comparan las muestras de los sitios desembocadura y testigo, estos grupos de muestras forman sus clases respectivas; esto significa que aún con la posible similitud manifestada en las figuras previas, las muestras de la clase B son las del sitio testigo y las muestras de la clase A son las del sitio desembocadura.

Con los resultados de las comparaciones observadas en la Figura 25, al realizar el comparativo entre clases, ninguna de ellas se mezcla con otra; sin embargo, algunas muestras se ubican en el límite de las fronteras de clases. También se puede observar que en este modelo la región de confianza de las clases es mayor en algunos gráficos que en otros, tanto en la clase A como en la clase B.

Estos resultados contribuyen a precisar que aún cuando las muestras del Gleysol, de los sitios de muestreo desembocadura y testigo, presentan similitud en sus propiedades físicas y químicas, con el uso de la tecnología NIRS se puede generar información útil y suficiente que nos permita establecer modelos de clasificación mediante técnicas quimiométricas (SIMCA).

6.4.1.4. Modelo de clasificación SIMCA para el Histosol soditerrico

La generación del modelo de clasificación SIMCA para el suelo Histosol soditerrico, se presenta en la Figura 26 y se realizó mediante un preprocesado de los datos de la escala de la varianza y una transformación aplicando una primer derivada. Este modelo se generó con tres sitios de muestreo.

Con el modelado de esta matriz de suelo se observa que en la Figura (26a) al hacer la comparación entre clases A y B, las muestras están en sus respectivas clases. En la clase A se encuentran muestras del sitio desembocadura, en la clase B se localizan las del sitio orilla de campo, y en la zona de muestras desconocidas, por no presentar características similares a las muestras de las clases A ó B, se encuentran las muestras pertenecientes al sitio de derrame. La región de confianza de cada clase es muy similar siendo amplia en cada caso.

En el gráfico (26b) se observa que al comparar la clase B, correspondiente a las muestras de suelo del sitio desembocadura, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. En el caso de la zona de las muestras desconocidas encontramos a las muestras orilla de campo las cuales se encuentran muy cercanas a la frontera de la clase B.

La región de confianza de la clase A es menor a la de la clase B (Figura (26b)), a pesar de esto el grupo de muestras se encuentra separado al límite de la frontera de la clase, y este grupo de muestras es más compacto que el de la clase A y que el de la zona de muestras desconocidas. No se presentan muestras en el cuadrante inferior izquierdo, lo que nos indica que no hay muestras que presenten similitud en sus características con las de la clase A ó B.

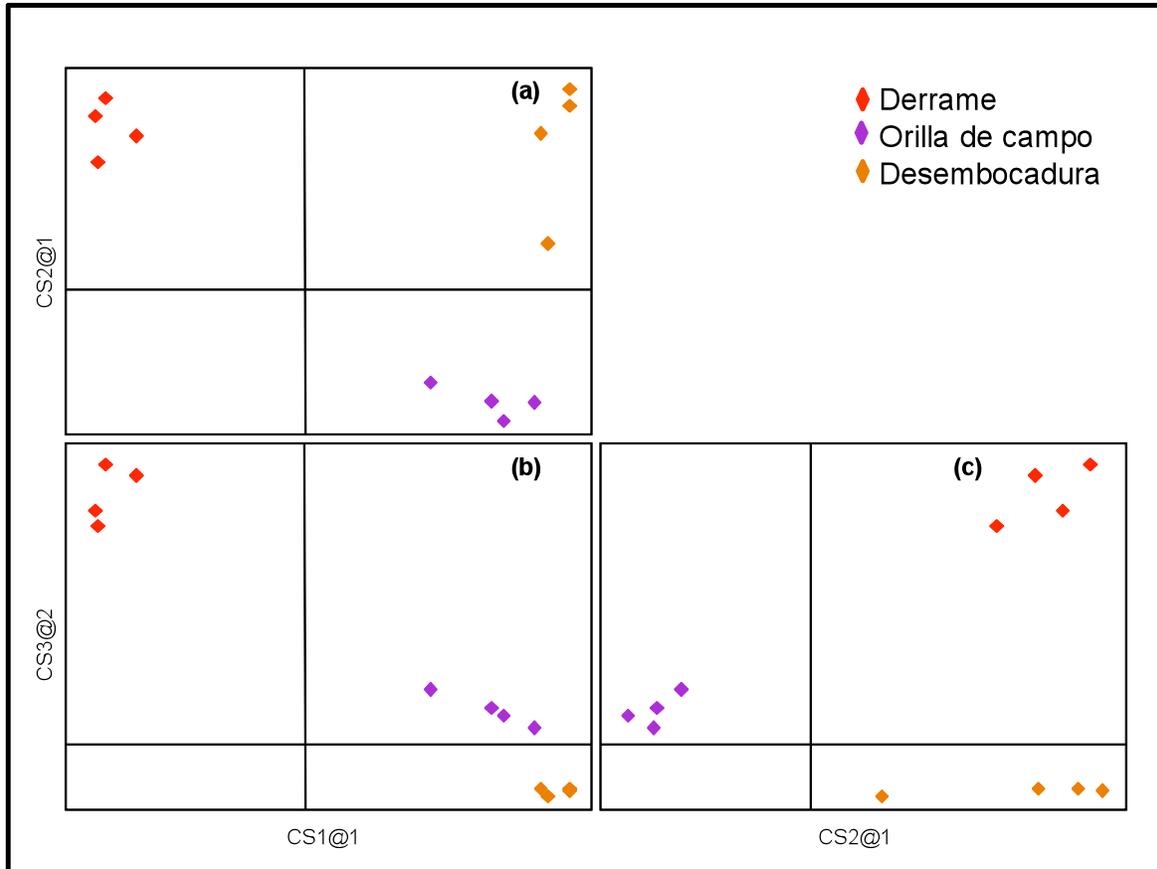


Figura 26. Diagrama de Coomans, construido con el modelado de clases SIMCA, para muestras del suelo Histosol soditerrico.

En la Figura (26c) se observa que al comparar la clase B, correspondiente a las muestras de suelo del sitio desembocadura, con la clase A, correspondiente a las muestras de suelo del sitio orilla de campo; cada una de estas se ubica en su respectiva área de acoplamiento. En el caso de las muestras desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A y B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas. Este grupo de muestras se encuentran alejadas de las fronteras de las clases A y B. No existen muestras con similitud a ambas clases.

Se puede observar que los grupos de muestras de las diferentes zonas de acoplamiento tienen diferente comportamiento siendo el más compacto el grupo de las muestras orilla de campo, seguido de las muestras del sitio derrame, y presentándose más separadas las muestras del sitio desembocadura. También observamos que la región de confianza de la clase A es mayor al de la clase B.

Con los resultados de las comparaciones observadas en la Figura 26, al realizar el comparativo entre clases, ninguna de ellas se mezcla con otra, ni se salen del área de acoplamiento; esto indica que la sensibilidad de los modelos son del 100%, dado que las cinco clases de muestras son incluidas en sus clases respectivas (A ó B) cuando son utilizadas para construir los modelos.

6.4.1.5. Modelo de clasificación SIMCA para el Histosol fibrico

Para el modelado de la matriz Histosol Fábriico (Figura 27) se realizó un preprocesado de los datos mediante la escala de rango y una transformación de los mismos, aplicando la primera derivada y calculando la variable normal estándar (SNV).

En la Figura (27a) se puede observar que al comparar la clase B, correspondiente a las muestras de suelo del sitio orilla del campo, con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en el caso de las muestras de las clases testigo y desembocadura, al no presentar semejanza con ninguna de las clases anteriores (clases A ó B) son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas.

Se observa que la región de confianza de la clase A, es ligeramente mayor a la clase B, a pesar de esto las muestras están lejanas a la frontera de cada clase. No existen muestras con similitud a ambas clases.

En la Figura 27**b**) se puede observar que en la clase B, se encuentran dos grupos de muestras, esto por que comparten ciertas características entre ellas, correspondiendo a las muestras de suelo del sitio desembocadura y testigo, y al compararlas con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Se puede observar una ligera dispersión de las muestras en la clase A, por lo cual una de las muestras del sitio testigo se encuentra cercana al límite de la frontera de la clase con respecto a la zona de muestras desconocidas. En el caso de las muestras de la clase orilla del campo, al no presentar semejanza con ninguna de las clases anteriores son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas, por no pertenecer a las clases A ó B. En esta zona es donde se muestra una mayor dispersión de las muestras con respecto a las otras clases. En el cuadrante inferior izquierdo, que corresponde a muestras con características similares a las de las clases A y B, no se encuentran muestras.

En la Figura (27**c**) se puede observar que en la clase B, se encuentran dos grupos de muestras, esto es por que comparten ciertas características entre ellas, correspondiendo a las muestras de suelo del sitio desembocadura y testigo, y al compararlas con la de clase A, correspondiente a las muestras de suelo del sitio orilla de campo; cada una de estas se ubica en su respectiva área de acoplamiento. Se puede observar una dispersión de las muestras en la clase A siendo mayor en esta clase que en las muestra de la clase B.

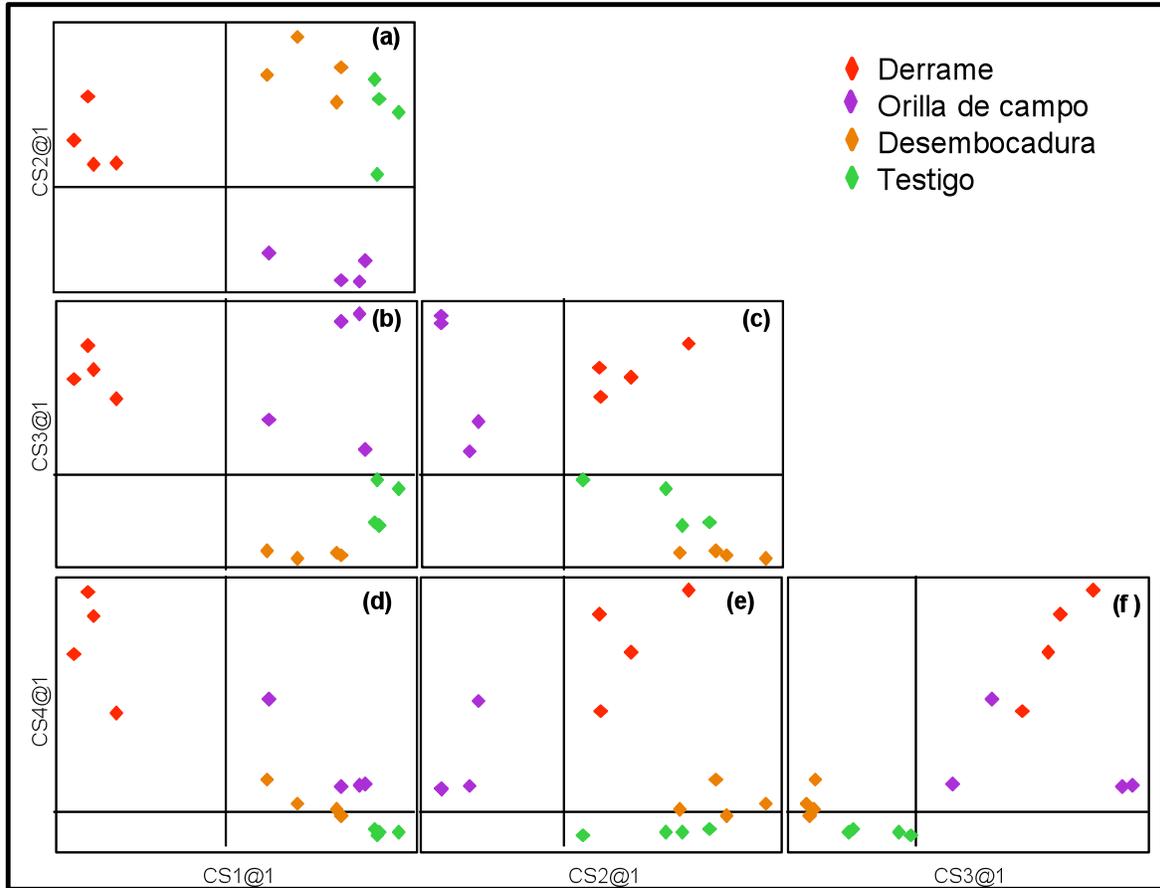


Figura 27. Diagrama de Coomans construido con el modelo de clases SIMCA, para muestras de suelo Histosol fibrico.

En la clase B (Figura (27c)) se puede observar que por la dispersión dada entre las muestras, las del sitio testigo se encuentran cercanas al límite de la frontera con la zona de muestras desconocidas y quedando más alejada de esta las muestras del sitio desembocadura.

En el caso de las muestras de la clase derrame, al no presentar semejanza con ninguna de las clases anteriores son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas, por no pertenecer a las clases A ó B. En el cuadrante inferior izquierdo, que corresponde a muestras con similitud en las características a las de las clases A y B, no se encuentran muestras.

En la Figura (27**d**) se observa que en la clase B, correspondiente a las muestras de suelo del sitio testigo comparte características con una muestra del grupo del sitio desembocadura, y al compararla con la clase A, correspondiente a las muestras de suelo del sitio derrame; cada una de estas se ubica en su respectiva área de acoplamiento. Y en la zona de las muestras desconocidas se encuentran muestras del sitio orilla de campo y testigo, esta última comparte una de las muestras con la clase A.

Se puede observar que existe similitud en las muestras desembocadura y testigo y que las regiones de confianza de cada clase son diferentes, siendo mayor la región de la clase A con respecto a la clase B.

En la Figura (27**e**) se observa que en la clase B, correspondiente a las muestras de suelo del sitio testigo comparte características con una muestra del grupo del sitio desembocadura, y al compararla con la clase A, correspondiente a las muestras de suelo del sitio orilla de campo cada una de estas se ubica en su respectiva área de acoplamiento. La región de confianza de ambas clases es diferente, siendo mayor en la clase A con respecto a la clase B.

En el caso de las muestras de las clases derrame y desembocadura, al no presentar semejanza con ninguna de las clases anteriores son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas, por no pertenecer a las clases A ó B. Las muestras del sitio desembocadura comparte una de las muestras con la clase A.

Por último, en la Figura (27**f**) se muestra que en la clase B, no se encuentra muestra, puesto que estas, a diferencia de la clase A en la cual se localizan las muestras del sitio desembocadura, sin embargo se encuentra en el límite de la frontera de esta clase compartiendo una de ellas en la zona donde se encuentran las muestras con características de las clases A y B.

En el caso de las muestras de las clases orilla del campo y derrame, al no presentar semejanza con ninguna de las clases anteriores son colocadas en el cuadrante superior derecho, que corresponde a muestras desconocidas, por no pertenecer a las clases A ó B. Se observa que en esta zona las muestras se encuentran más dispersas a diferencia de las muestras de las otras clases.

Estos resultados contribuyen a precisar que aún cuando las muestras del Histosol fibrico, en los sitios de muestreo desembocadura y testigo, presentan similitud en sus propiedades físicas y químicas, el uso de la tecnología NIRS sirve para generar información útil y suficiente que nos permita establecer modelos de clasificación mediante técnicas quimiométricas (SIMCA).

Con los resultados de las comparaciones observadas en la Figura 27, al realizar el comparativo entre clases, se observa que las muestras del sitio desembocadura y testigo resultaron similares, y se salen del área de acoplamiento, es decir, que la sensibilidad de los modelos son menores al 100%; dado que las cinco clases de muestras son incluidas en las clases respectivas (A ó B) cuando son utilizadas para construir los modelos.

6.4.2. Distancia de clases

En la Figura 28 se muestra la distancia de clases, proyección de clases o grado de separación de las clases; esto permite tener una evaluación del grado de separación de las clases o categorías. En el Histosol soditérico, no se tiene testigo, este fue el mismo que para el Histosol fibrico. En todos los casos las nubes de puntos presentan una separación adecuada para observar una buena distancia de clases entre las categorías establecidas durante el muestreo. En el Solonchack se presenta una excepción debido a

que las muestras de los sitios testigo y desembocadura están en las mismas coordenadas (valores de los componentes principales). Las muestras del sitio derrame siempre presentaron una mayor distancia con respecto a las otras categorías. Las distancias de clase que se presentan pueden atribuirse al grado de contaminación por hidrocarburos derivados del petróleo.

6.4.3. Predicción de clases

En el Cuadro 11 se muestra la predicción de clases en función del tipo de suelos y del sitio de muestreo (a los cuales se les asignó un número que representa una clase respectivamente). Los datos indican que para los cinco suelos es posible obtener una predicción de clases del 100 %, en virtud de que cada muestra es clasificada en su respectiva categoría y que solo se presentan los resultados de esta prueba en una sola columna. Por tanto, las muestras de los sitios de muestreo derrame, orilla, desembocadura y testigo son perfectamente clasificadas en su clase respectiva.

En esta etapa de la investigación se comprobó que para establecer modelos de clasificación de suelos contaminados la tecnología NIRS proporciona información suficiente.

Durante la interpretación quimiométrica fue necesario reparar las huellas espectrales en función del tipo de suelo para el establecimiento de los modelos.

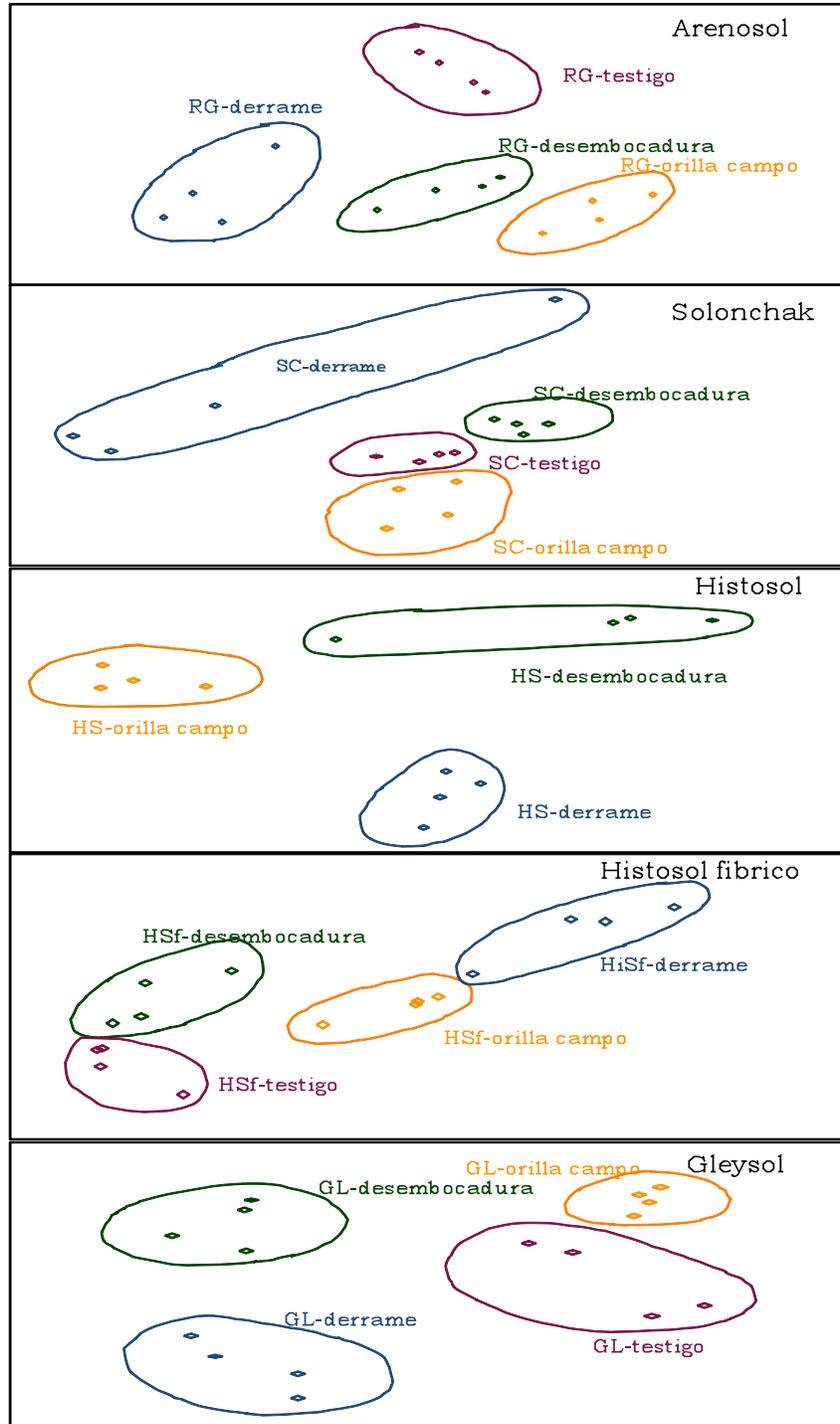


Figura 28. Proyección de clases en el modelo SIMCA para las categorías de los cinco suelos.

Cuadro 11. Predicción de clases de las muestras de suelo, en función del sitio de muestreo.

Sitio de muestreo	Suelo				
	Gleysol	Solonchack	Histosol soditérico	Histosol fibrico	Arenosol
	Predicción de clase				
Derrame	1	1	1	1	1
Derrame	1	1	1	1	1
Derrame	1	1	1	1	1
Derrame	1	1	1	1	1
Orilla	2	2	2	2	2
Orilla	2	2	2	2	2
Orilla	2	2	2	2	2
Orilla	2	2	2	2	2
Desembocadura	3	3	3	3	3
Desembocadura	3	3	3	3	3
Desembocadura	3	3	3	3	3
Desembocadura	3	3	3	3	3
Testigo	4	4	4	4	4
Testigo	4	4	4	4	4
Testigo	4	4	4	4	4
Testigo	4	4	4	4	4

6.5. Contenido de Hidrocarburos Totales en las Muestras de Suelo

El grado de contaminación de los suelos de la Cuenca baja del río Tonalá, por hidrocarburos totales del petróleo (TPH), según las determinaciones realizadas mediante el método Soxhlet y el análisis de varianza realizado demuestra que se encuentra en diferentes niveles de contaminación, esto conforme a sitios de muestreos desde el campo petrolero hacia la desembocadura de la cuenca y dependiendo del tipo de suelo. En el Cuadro

12 se presentan los valores promedio de las muestras de hidrocarburos totales.

Se observa que la concentración de hidrocarburos derivados del petróleo, en las muestras de suelo, es significativamente mayor en el sitio del derrame que en el resto de los puntos de muestreo. La tendencia en la concentración va de mayor a menor, en el siguiente orden: sitio del derrame, orilla de campo, desembocadura y testigo. Los contenidos de los hidrocarburos totales del petróleo, en el sitio del derrame, se encuentran en el intervalo de 4,679 (Arenosol) a 419,198 (Solonchack) mg kg^{-1} de TPH en suelo seco.

La concentración de hidrocarburos totales en las muestras es estadísticamente mayor en los sitio del derrame, en comparación de los sitios restantes, en cada uno de los suelos analizados, mientras que en los sitios de muestreo orilla de campo, desembocadura y testigo de cada suelo, no se encontraron diferencias estadísticas significativas.

Con base en lo establecido por la EPA (1996), en el método 4030, el 74 % de las muestras presentaron un nivel de concentración de hidrocarburos totales (TPH) considerada como alto (superior a $1,000 \text{ mg kg}^{-1}$ de TPH en suelo seco) y el 26 % restante presentó una concentración considerada como media ($>100 \text{ mg kg}^{-1}$ de TPH en suelo seco y $<1,000 \text{ mg kg}^{-1}$ de TPH en suelo seco).

A continuación se presenta un análisis de la contaminación en cada tipo de suelo, lo cual deriva del objetivo específico 2.1.1.3.

6.5.1. Histosol soditérico

Para el suelo Histosol sódico, se puede observar que el grado de contaminación presenta un comportamiento según el transecto de muestreo (Cuadro 12), con un grado de contaminación mayor en el sitio de derrame

(98,070 mg kg⁻¹ de TPH en suelo seco) en el campo petrolero Cinco Presidentes, disminuyendo conforme se aleja del sitio del derrame encontrando en la zona de la orilla del campo petrolero una concentración de 17,012 mg kg⁻¹ de TPH en ss; mientras que en la desembocadura de la cuenca presenta una concentración de 15,172 mg kg⁻¹ de TPH en ss. En este suelo no se incluyó un sitio testigo. Por las concentraciones cuantificadas de TPH, en todos los sitios de muestreo del Histosol sódico, según la EPA, el grado de contaminación de este suelo se encuentra en un nivel alto, debido a que superan los 1,000 mg kg⁻¹ de TPH en suelo seco, sin embargo, el sitio con derrame registró una concentración de TPH siete veces superior a los sitios de orilla y desembocadura, lo cual se atribuye a hidrocarburos aportados por un derrame de petróleo ubicado a 50 m, procedente de otro pozo petrolero.

Estos resultados son superiores a los reportados por García *et al.* (2006), quienes encontraron una concentración máxima de 1,708 mg kg⁻¹ ss, por lo cual se deduce que este sitio de muestreo presenta una contaminación crónica por hidrocarburos derivados del petróleo.

Cuadro 12. Contenido promedio de hidrocarburos totales en las muestras de suelo.

Suelo y procedencia	mg kg ⁻¹ de TPH en suelo seco
Histosol sódico	
HS1 (sitio del derrame)	98,070 a
HS2 (orilla de campo)	17,012 b
HS3 (desembocadura)	15,172 b
Histosol fibrico	
HSf1 (sitio del derrame)	66,320 a
HSf2 (orilla de campo)	21,330 b
HSf3 (desembocadura)	16,428 b
HSf4 (testigo)	16,861 b
Solonchak	
SC1 (sitio del derrame)	419,198 a
SC2 (orilla de campo)	4,535 b
SC3 (desembocadura)	13,635 b
SC4 (testigo)	6,560 b
Gleysol	
GL1 (sitio del derrame)	6,874 a
GL2 (orilla de campo)	1,493 b
GL3 (desembocadura)	793 b
GL4 (testigo)	857 b
Arenosol	
RG1 (sitio del derrame)	4,679 a
RG2 (orilla de campo)	464 b
RG3 (desembocadura)	252 b
RG4 (testigo)	201 b

TPH=hidrocarburos totales del petróleo; medias con la misma letra en columna por tipo de suelo no son estadísticamente diferentes (Tukey, p=0.05)

6.5.2. Histosol fibrico

En el Histosol fibrico la cuantificación realizada, en las muestras procedentes del sitio del derrame, presentaron una concentración de 66,320 mg kg⁻¹ TPH en suelo seco del campo petrolero Sánchez Magallanes, siendo tres a cuatro veces superior a las concentraciones encontradas en los sitios orilla de campo, desembocadura y testigo (Cuadro 12). El sitios de muestreo que presento menores contenidos de TPH fue el desembocadura con una concentración de 16,428 mg kg⁻¹ de TPH en ss. Estas concentraciones, son clasificadas como un nivel de contaminación alta, de acuerdo con la EPA (1996) en el método 4030.

Estos resultados, comparados con los reportados por Zavala *et al.*, 2005, son superiores lo que indica que en este sitio de muestreo se presenta una afectación crónica en cuanto a los contenidos de hidrocarburos derivados del petróleo.

Se puede observar que en el sitio de muestreo testigo la concentración de TPH es ligeramente mayor con respecto al sitio de la desembocadura de la cuenca, esto puede suceder debido a que el material orgánico que presenta este tipo de suelo, influye en las concentración de compuestos constituidos por carbono (biogenicos) (Calva *et al.*, 2005), ya que las muestras de este sitio fueron tomadas alejadas de la desembocadura de la cuenca y de campos petroleros.

6.5.3. Solonchak

En el suelo Solonchack las muestras procedentes del sitio del derrame presentan un grado de contaminación promedio de 419,198 mg kg⁻¹ de TPH en suelo seco seguido del sitio desembocadura con 13,635 mg kg⁻¹ de TPH

en suelo seco; siendo las muestras de suelo procedentes de la orilla del derrame las que presentan nivel de hidrocarburos más bajos (4,535 mg kg⁻¹ de TPH en suelo seco) (Cuadro 12). Estos resultados indican que el suelo con derrame de petróleo procedente de la petroquímica La Venta tiene entre 91 y 31 veces más hidrocarburos en comparación con los suelos de desembocadura, orilla de campo y testigo. En todos los sitios la concentración de hidrocarburos es superior a los permitidos por las normas EPA, es decir superan el límite máximo permisible de 1,000 mg kg⁻¹ de TPH en suelo seco.

6.5.4. Gleysol

Las muestras del suelo Gleysol presentaron un grado de contaminación promedio (Cuadro 12), en el sitio del derrame de 6,874 mg kg⁻¹ de TPH en suelo seco en el campo petrolero Tucán, seguido del sitio de orilla de campo con 1,493 mg kg⁻¹ de TPH en suelo seco, siendo aproximadamente cinco veces mayor la contaminación del sitio de derrame con respecto a la orilla de campo. En este suelo al comparar la los niveles de concentración encontrados se tiene una diferencia de ocho veces superior la contaminación del sitio del derrame con respecto a las muestras del sitio con menor grado de contaminación.

De acuerdo con las normas de la EPA estos suelos presentan niveles de contaminación alta. En el caso de las muestras de los sitios desembocadura de la cuenca y el sitio testigo presentan concentraciones inferiores a 1,000 mg kg⁻¹ de TPH por tanto, estos dos sitios de muestreo son clasificados con base en las normas EPA como suelos con niveles de hidrocarburos medios (menor a 1,000 mg kg⁻¹ pero superiores a 100 mg kg⁻¹).

Estos resultados comparados con los obtenidos por Zavala *et al*, 2005 son inferiores puesto que estos autores reportan concentraciones superiores a los 14, 000 mg kg⁻¹ de HTP en los sitios del derrame.

6.5.5. Arenosol

En el suelo Arenosol se cuantificó un nivel de contaminación promedio, en el sitio del derrame de 4,679 mg kg⁻¹ de TPH en suelo seco (Cuadro 12), y se observa que el comportamiento en los niveles de contaminación por hidrocarburos fue según el transecto de muestreo descendiendo desde el sitio de derrame hasta el testigo. Los sitios orilla de campo, desembocadura y testigo, se observa que están en un intervalo de 464 a 200 mg kg⁻¹ de TPH en suelo seco, siendo estos sitios de muestreo 10 y 23 veces inferiores en las concentraciones de hidrocarburos con respecto al sitio del derrame en los cuales el nivel de contaminación se clasificó como medio con base en las normas de la EPA.

Al comparar estos resultados con los reportados con otros autores como es el caso de Adams *et al.*, 2008 quien reporta una concentración de 2, 200 mg kg⁻¹ de HTP y por Zavala *et al.*, 2005 que reporta concentraciones superiores a los 8, 000 mg kg⁻¹ de THP, este sitio de muestreo va disminuyendo sus concentraciones de hidrocarburos del petróleo con el paso de los años.

En el Cuadro 12 se observa que el grado de contaminación puede estar determinado por el tipo de suelo, debido a que los suelos con mayor contenido de materia orgánica (Histosol) y más arcilla (Solonchack) son los que presentan más altos contenidos de hidrocarburos. Lo cual puede atribuirse a que estos materiales retienen más hidrocarburos que los suelos con poca materia orgánica y poca arcilla (Arenosol y Gleysol) (Li *et al.*, 1997; Nava y Tirado, 1980)

En el caso de los Histosoles estos suelos permanecen más tiempo saturados de agua, lo cual impide la aireación del petróleo derramado ya que este se puede encontrar de forma libre o en emulsión con el agua y como consecuencia la degradación es más lenta (Reeves, 2005). Además de que la descomposición de la materia orgánica puede aportar hidrocarburos biogénicos que son extraídos junto con los hidrocarburos derivados del petróleo.

Se puede observar que si se compara el grado de contaminación de suelos en el sitio del derrame, el suelo Solonchak es el que presenta los niveles más altos de TPH, con respecto a los otros suelos, seguido de las muestras del suelo Histosol, Gleysol y Arenosol respectivamente.

6.6. Modelo para Estimación de Hidrocarburos Totales en Suelos Mediante NIRS

Con objeto de evaluar la capacidad de la espectroscopia del infrarrojo cercano para cuantificar los contenidos de hidrocarburos derivados del petróleo en muestras de suelo, se procedió a generar un modelo de calibración y estimación con base en la señal espectral y los datos obtenidos con el método de referencia.

Debido a que en los modelos de clasificación se encontró que existe una fuerte influencia de la matriz de suelos, para clasificar y predecir por medio de la tecnología NIRS, en este apartado se buscó encontrar la tendencia que presenta el modelo para la estimación de los contenidos de hidrocarburos en suelos obteniéndolos con NIRS.

Para el desarrollo del modelo se analizaron diferentes ecuaciones de calibración. Entre los tratamientos matemáticos aplicados se incluyeron:

correcciones de los espectros por problemas ópticos sin modificar la frecuencia (derivadas), correcciones por tamaño de partícula (Gap) y correcciones de la tendencia de los datos (Smooth). Para indicar el tratamiento aplicado se indica en el primer número del modelo matemático, el orden de la derivación o substracción (uno es la primera derivada del logaritmo $1/R$), el segundo número es la amplitud o distancia entre los segmentos a sustraer (Gap), el tercero son puntos suavizadores (Smooth) en cada segmento y el cuarto número son los segundos puntos suavizadores (Smooth 2).

En el Cuadro 13 se mencionan las ecuaciones creadas a partir de diferentes modelos matemáticos probados, con sus respectivos estadísticos. En el se observa que el promedio de TPH de las 63 muestras de suelos en las ecuaciones probadas tuvo un rango de entre 13,285 y 34,118 mg kg⁻¹. La menor desviación estándar (SD) la presenta la ecuación 4 con 21,726 mg kg⁻¹ en suelo seco, el menor error estándar de calibración (SEC) o varianza de la ecuación, lo presenta la ecuación 11 con un valor de 2,647, seguida por la ecuación 12 y 13 con 3,140 mg kg⁻¹ en suelo seco de TPH. El coeficiente de determinación (RSQ) de las 16 ecuaciones tuvo un intervalo entre 0.189 y 0.988, mostrando buenos ajustes de los modelos, sin embargo la ecuación 11 fue la que presento el mayor valor para esta variable.

La menor varianza de los puntos redundantes (SECV) la presenta la ecuación 4 con un valor de 5196. La varianza residual debe tender a 1, y es la ecuación 10 la que estuvo más cerca de la unidad con un valor de 0.95. En general a pesar de que las ecuaciones presentan un error estándar de calibración alto, esto debido a la heterogeneidad en los niveles de los contenidos de TPH y alta diferencia entre los mismos, se observa que el coeficiente de determinación es alto en varias de las ecuaciones de calibración

Cuadro 13. Validación estadística de las ecuaciones de calibración para la validación del modelo de estimación de HTP mediante NIRS.

Modelo y rango	Tratamiento matemático	Media	SD	SEC	RSQ	SECV	1-VR
Ecuación 01	sin correctivo Derivada 0, Gap 0, Smooth 1:1	34118	91702	82580	0.189	87697	0.07
Ecuación 02	sin correctivo Derivada 1, Gap 4, Smooth 4:1	34118	91702	55150	0.638	62958	0.52
Ecuación 03	sin correctivo Derivada 2, Gap 4, Smooth 4:1	34118	91702	10200	0.988	45555	0.76
Ecuación 04	sin correctivo Derivada 2, Gap 10, Smooth 10:1	13285	21726	3234	0.978	5196	0.94
Ecuación 05	SNV Derivada 0, Gap 0, Smooth 1:1	15531	24541	10530	0.816	14158	0.67
Ecuación 06	SNV Derivada 1, Gap 4, Smooth 4:1	15339	24881	5392	0.953	8704	0.88
Ecuación 07	SNV Derivada 2, Gap 4, Smooth 4:1	15303	24530	3666	0.978	10229	0.83
Ecuación 08	SNV Derivada 2, Gap 10, Smooth 10:1	17185	27451	5537	0.959	8642	0.90
Ecuación 09	DET Derivada 0, Gap 0, Smooth 1:1	16180	26573	6792	0.935	8030	0.91
Ecuación 10	DET Derivada 1, Gap 4, Smooth 4:1	15792	26643	3527	0.983	5674	0.95
Ecuación 11	DET Derivada 2, Gap 4, Smooth 4:1	14908	24772	2647	0.989	6410	0.93
Ecuación 12	DET Derivada 2, Gap 10, Smooth 10:1	14354	24010	3140	0.983	5363	0.95
Ecuación 13	MSC Derivada 0, Gap 0, Smooth 1:1	14354	24010	3140	0.983	5363	0.95
Ecuación 14	MSC Derivada 1, Gap 4, Smooth 4:1	13893	21755	6672	0.906	10902	0.75
Ecuación 15	MSC Derivada 2, Gap 4, Smooth 4:1	14907	24553	5807	0.944	12045	0.76
Ecuación 16	MSC Derivada 2, Gap 10, Smooth 10:1	18513	29846	6304	0.955	10482	0.88

SD= desviación estándar; SEC= error estándar de calibración; RSQ=coeficiente de correlación múltiple; SECV= error estándar de validación cruzada; 1-VR= 1-VR estimación de la R²

6.6.1. Validación interna de las ecuaciones

Se realizó una validación interna de la ecuación 3 mediante validación cruzada (predicción del contenido de TPH en una muestra y construyendo el modelo con el resto de muestras, y así hasta que todas las muestras son estimadas) (Figura 29). En esta validación interna del modelo de predicción se obtuvo un RSQ de 0.989, y el error estándar en la predicción (SEP) encontrado es de 9443.260. Así mismo el error estándar de predicción en la calibración (SEP (C)) es de 9519.110 y el BIAS encontrado fue de 0.010, con esto se muestra que el modelo presenta buena tendencia para la predicción de HTP mediante NIRS

En la misma Figura 29 se puede observar que el agrupamiento de las muestras se encuentra en la base de la línea, esto debido a que PLS se fundamenta en el agrupamiento de componentes principales. En la figura se observan dos muestras en el otro extremo de la línea de tendencia, estas son las que tienen mayor contenido de hidrocarburos, con base en lo cuantificado con el método de referencia (extracción Soxhlet).

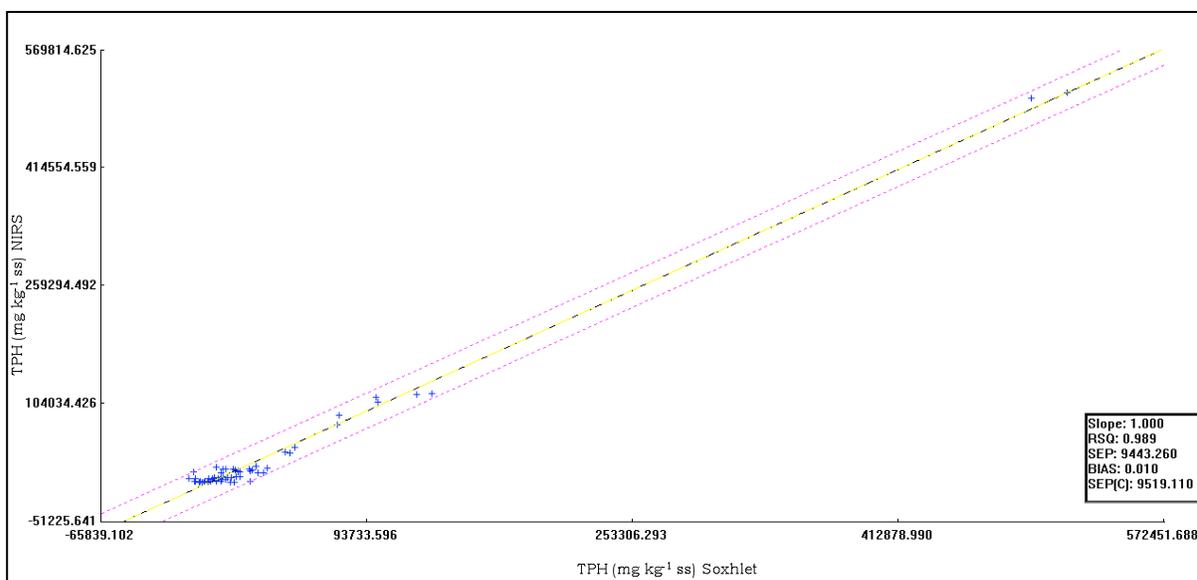


Figura 29. Modelo de validación interna para estimar TPH mediante NIRS.

En el Cuadro 14 podemos observar que para la estimación y predicción de los valores mediante NIRS, al comparar los contenidos de hidrocarburos totales del petróleo que se obtuvieron en Soxhlet con los que predice NIRS, se presentaron muestras que de forma individual se comportaron errantes mostrando que hay una gran diferencia entre ellos entre un 15% hasta más del 100%, proporcionando datos de predicción erróneos, al grado que no pueden ser considerados como errores aceptables. Esto se presenta debido a la heterogeneidad en los contenidos de hidrocarburos que se presentan en las muestras puesto que se están usando muestras que tienen niveles por debajo de los 1,000 mg kg⁻¹ y muestras que están por arriba de los 5,000 mg kg⁻¹, otra causa es la diferencia en las matrices de suelo usadas. No obstante lo anterior, presenta una tendencia aceptable en la estimación de hidrocarburos, en suelos mediante la señal espectral obtenida con la tecnología NIR en cada muestra de suelo.

Cuadro 14. Constituyentes estadísticos del modelo de predicción interna.

No. Sample	Soxhelet	NIRS	Residual	Bias
s02	116721.30	132662.64	-15941.344	-15941.35
s03	75467.10	75715.47	-248.367	-248.38
s04	88691.80	76803.21	11888.586	11888.58
s05	111403.50	99403.25	12000.250	12000.24
s07	17730.70	13539.21	4191.492	4191.48
s08	14197.20	23564.75	-9367.554	-9367.56
s10	21386.90	27156.13	-5769.225	-5769.24
s11	13236.20	16256.33	-3020.133	-3020.14
s12	13787.90	17601.89	-3813.988	-3814.00
s13	12277.70	31363.59	-19085.895	-19085.90
s15	114994.00	123454.54	-8460.539	-8460.55
s16	38294.20	47713.14	-9418.941	-9418.95
s17	39119.50	44782.33	-5662.828	-5662.84
s18	5207.20	-1196.35	6403.545	6403.54
s19	19575.00	3039.81	16535.191	16535.18
s20	45665.40	50399.77	-4734.367	-4734.38
s21	14875.20	16377.78	-1502.578	-1502.59
s22	11877.50	28281.99	-16404.492	-16404.50
s24	17743.50	6975.79	10767.707	10767.70
s25	18241.60	6975.79	-15796.600	-15796.61
s26	15724.60	24757.36	-9032.758	-9032.77
s27	17663.40	8902.65	8760.751	8760.74
s28	17214.70	13415.81	3798.893	3798.88
s29	16841.60	23171.76	-6330.156	-6330.17
s31	513356.41	514425.25	-1068.844	-1068.85
s32	105764.10	100300.06	5464.039	5464.03
s33	506775.09	492826.56	13948.531	13948.52
s35	4801.90	1067.24	3734.659	3734.65
s36	4293.40	-13087.80	17381.197	17381.19
s37	3984.20	8590.31	-4606.109	-4606.12
s34	5061.90	4026.19	1035.712	1035.70
s38	11987.10	6132.08	5855.024	5855.01
s39	13438.90	-10651.44	24090.342	24090.33
s40	15790.00	14351.76	1438.243	1438.23
s41	13325.60	17274.28	-3948.678	-3948.69
s42	6036.70	9920.58	-3883.883	-3883.89

Cuadro 14. Constituyentes estadísticos del modelo de predicción interna. (Continuación).

s43	6234.20	12439.50	-6205.299	-6205.31
s45	6720.40	17631.86	-10911.459	-10911.47
s46	7079.10	14822.74	-7743.640	-7743.65
s47	5681.20	1884.78	3796.420	3796.41
s48	7788.20	7238.01	550.188	550.18
s49	6946.80	17623.05	-10676.253	-10676.26
s50	1276.80	-9821.99	11098.789	11098.78
s51	3948.30	1047.51	2900.795	2900.79
s52	640.00	-9563.74	10203.739	10203.73
s54	741.20	2968.35	-2227.147	-2227.16
s55	211.10	11692.20	-11481.100	-11481.11
s56	636.00	-1539.36	2175.363	2175.35
s57	1585.90	-3620.79	5206.693	5206.68
s58	535.00	-9477.76	10012.756	10012.75
s59	963.50	-113.93	1077.425	1077.42
s61	1502.70	23740.14	-22237.443	-22237.45
s62	4653.60	-9458.98	14112.584	14112.57
s63	2420.20	2981.73	-561.528	-561.54
s64	4043.00	6462.71	-2419.714	-2419.72
s66	506.50	-2171.51	2678.008	2678.00
s67	710.30	6090.84	-5380.542	-5380.55
s69	640.10	-6541.77	7181.866	7181.86
s70	100.80	-6928.50	7029.294	7029.28
s71	403.30	-5541.00	5944.304	5944.29
s72	302.90	13881.12	-13578.218	-13578.23
s73	201.80	-5457.62	5659.416	5659.41
s74	402.10	-4196.36	4598.462	4598.45

6.6.2. Validación externa del modelo de predicción

De las ecuaciones iniciales se seleccionó la ecuación 3 para el desarrollo de la validación externa, por ser la que presentó los mejores parámetros estadísticos (RSQ, SEC y SEP).

En la validación se utilizó un archivo de muestras que no fueron consideradas para generar los modelos de calibración, este archivo contiene 10 muestras las cuales se les denominan como extrañas al modelo, estas fueron registradas igual que las del grupo de calibración, posteriormente se utilizó el modelo elegido y se predijo el contenido de TPH.

Estadísticamente es recomendable utilizar un mínimo de 30 datos, para construir el modelo con la finalidad de que estos sean normales y presenten tendencia en el agrupamiento de las muestras; en este estudio solo se pretende ver la viabilidad del modelo, obtenido con la señal espectral del NIRS, para estimar la concentración de TPH en suelos.

En la validación externa del modelo de predicción se observó que la estimación de TPH es aceptable (Figura 30) debido a que en los parámetros estadísticos analizados se encontró un RSQ de 0.970, el error estándar de predicción (SEP) es de 42,165.820, el error estándar en la calibración (SEC) es de 43,530.223 y el BIAS encontrado es de -8518.458.

En este modelo se observa que aún cuando el número de muestras es bajo, la mayoría de ellas presentan una tendencia a agruparse en la base del gráfico donde se encuentran los contenidos más bajos de hidrocarburos, esto porque en sus contenidos de hidrocarburos presentan valores en un intervalo similar entre ellas a excepción de tres muestras, las cuales en sus contenidos de hidrocarburos presentan una variación alta, por lo cual se separan del grupo y de los límites de la tendencia.

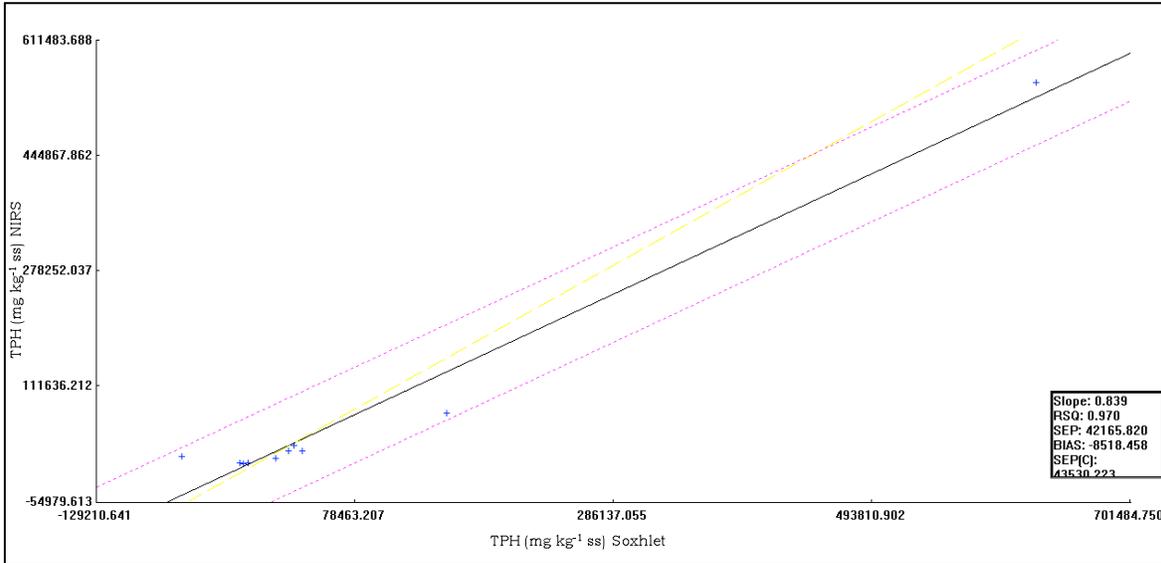


Figura 30. Modelo de validación externa para estimar TPH mediante NIRS.

Lo discutido en el párrafo anterior permite inferir que cuando se validan los modelos de estimación con muestras extrañas al modelo, se debe considerar el rango de concentración probable de las mismas, así como un número alto de las muestras para reducir la varianza entre ellas. Se debe crear un modelo para cada unidad de suelo, por el efecto matricial que se detecto al interpretar las huellas espectrales.

CONCLUSIONES

Con base en la hipótesis general, la cual plantea que “la información generada mediante la tecnología NIRS es suficiente para clasificar y cuantificar la contaminación por hidrocarburos derivados del petróleo en suelos agrícolas con resultados similares a los obtenidos mediante métodos convencionales” se concluye que la tecnología NIRS con apoyo de la quimiometría permite obtener modelos para clasificar y estimar la contaminación de hidrocarburos en suelos, con resultados que superan la sensibilidad del método de referencia (extracción Soxhlet).

En el caso de la hipótesis específica 3.1.1, la cual plantea que “mediante la tecnología NIRS es posible generar modelos que puedan identificar suelos agrícolas contaminados con hidrocarburos derivados del petróleo, similar a los resultados obtenidos con los métodos convencionales” se concluye que las huellas espectrales generadas en cada muestra de suelo mediante NIRS, son suficientes para establecer modelos de clasificación con más sensibilidad para discriminar las muestras de suelos en función de su origen, que lo obtenido con el método de referencia (extracción Soxhlet).

Para que la hipótesis 3.1.1, sea considerada válida es necesario realizar un tratamiento quimiométrico (preprocesados y transformaciones) a los espectros NIRS y realizar la interpretación de las huellas espectros de las muestras de suelos en agrupamientos definidos por la unidad de suelo (Arenosol, Gleysol, Solonchack, Histosol, etc.), a fin de eliminar el efecto matricial que genera cada grupo de suelo; concretamente por el contenido de arcilla y materia orgánica.

Para la hipótesis específica 3.1.2, la cual plantea que “mediante la tecnología NIRS es posible generar modelos para cuantificar los contenidos de hidrocarburos derivados del petróleo en suelos agrícolas, con resultados similares a los obtenidos con los métodos convencionales” se concluye que los espectros NIRS puede ser información suficiente para establecer modelos de estimación de la concentración de hidrocarburos totales del petróleo cuyos resultados sean similares a los obtenidos por el método convencional de referencia. Sin embargo, deberán tenerse las siguientes consideraciones: a) cubrir con muestras suficientes el intervalo de estimación del modelo, b) aumentar el número de muestras, y c) establecer un modelo de estimación para cada grupo de suelo.

LITERATURA CITADA

- 1.- Adams, S. R. H., Domínguez, R. V. y García H. L. 1996. Potencial de la biorremediación de suelo y agua impactados por petróleo en el trópico mexicano. *Terra* 17:159-174
- 2.- Adams R.H, Guzman O.F.J, Zavala C.J. 2008. Water repellency in oil contaminated sandy and clayey soils. *J. Environ. Sci. Tech.* 5 (4), 445-454.
- 3.- Alomar D. 1998. Fundamentos de la espectroscopia de reflectancia en el infrarrojo cercano (NIRS) como método de análisis de forraje. *Agro Sur*, 26, 88-114.
- 4.- Analytical Methods Committee. 1994. *Analyst* pag 119
- 5.- Blanco M. y Cerdá V. 1988. *Quimiometría*. Publicaciones de la Universidad Autónoma de Barcelona. Bellaterra, España.
- 6.- Botello, V. A. 1996. Características, composición y propiedades fisicoquímicas del petróleo. *In* A. V. Botello, L. J. Rojas G., A. Benítez J. y L. D. Zárate (eds.). Golfo de México. Contaminación e impacto ambiental: Diagnostico y Tendencias. Programa EPOMEX. Universidad Autónoma de Campeche, Méx. pp. 204-210
- 7.- Brereton R. G. 1990. *Chemometrics Applications of Mathematics and Statistics to Laboratory Systems*. Ellis Horwood, Chichester, 489 pp.
- 8.- Brereton R. G. 1992. *Chemometrics: Applications of mathematics and chemometrics, Ilustred by Cases studies*. Amsterdam.

- 9.- Brereton R. G. 2000. Introduction to multivariate calibration in analytical chemistry. *Analyst* 125, 2125-2154.
- 10.- Brereton R.G., 2002. Time Series in Environmental Monitoring Handbook (F.R.Burden, I.McKelvie, U.Forster, A.Guenther, editors), McGraw Hill, Chapter 27 (p 27.1 – 27.16)
- 11.- Brereton R. G. 2003. Chemometrics Data Analysis for the Laboratory and Chemical Plant. Wiley, Chichester.
- 12.- Bouveresse E. 1997. Maintenance and Transfer of Multivariate Calibration Models Based on Near-Infrared Spectroscopy, Tesis doctoral, Vrije Universiteit Brussel. Bruselas, Belgica.
- 13.- Calva B.L.G., Botello V.A y Ponce V.G. 2005. Composición de hidrocarburos alifáticos en sedimentos de la laguna de Sontecomapan, Ver, México. *In: Hidrobiologica* 15(1):97-108.
- 14.- Camel V. 2002. Extraction techniques. *Analytical and Bioanalytical Chemistry* 372:39-40
- 15.- Carvalho A.R., Nogal S. M, J. Wattoon, and R. G. Brereton. 2006. Comparison of PLS and kinetic models for a second-order reaction as monitored using ultraviolet visible and mid-infrared spectrometry. *Talanta* 68: 1190-1200.
- 16.- Chang, C.W., Laird, D.A., 2002. Near-infrared reflectance spectroscopy analysis of soil C and N. *Soil Science*, 167(2):110-116.
- 17.- Chodak M., Nikli ska M. and Beese F. 2007. Near-infrared spectroscopy for analysis of chemical and microbiological properties of forest soil organic horizons in a heavy-metal-polluted area. *Biol Fertil Soils* 44: 171-180

- 18.- Confalonieri, M., Fornasier, F., Ursino, A., Boccardi, F., Pintus, B., Odoardi, M., 2001. The potential of near infrared reflectance spectroscopy as a tool for the chemical characterization of agricultural soils. *J. Near Infrared Spectrosc.* 9:123-131.
- 20.- Coomans D., and Broeckaert, I. 1986. *Potential Pattern Recognition in Chemical and Medical Decision Making*, Research Studies Press, Letchworth. 57 (3):233-247.
- 21.- Cozzolino D. 2002. Uso de la espectroscopia de reflectancia en el infrarrojo cercano (NIRS) en el análisis de alimento para animales. *Agrociencia* 6 (2) 24-32.
- 22.- Cozzolino D., Moron A. 2004. Exploring the use of near infrared reflectance spectroscopy (NIRS) to predict trace minerals in legumes. *In: Animal Feed Science and Technology* 111.
- 23.- Cozzolino D., Moron A. 2006. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. *In: Soil & Tillage Research* 85:78-85.
- 24.- Da Costa Filho P.A. 2008. Rapid determination of sucrose in chocolate mass using near infrared spectroscopy. *Analytica Chimica Acta* 631: 206-211.
- 25.- Dean, J. R. 1998. *Extraction methods for environmental analysis*. Johns Wiley & Sons. Chichester, England, ISBN 0-471-98287-3.
- 26.- Dorn, B. P., Vipond, E. T., Salanitro, P. J., y Wisniewski, L. H. 1998. Assessment of the acute toxicity of crude oils in soils using earthworms, microtox, and plants. *Chemosphere* 37: 845-860

- 27.- EPA 4030. 1996. Soil Screening for petroleum hydrocarbons by immunoassay. (Revisión 0, Diciembre de 1996).
- 28.- EPA 3540 C. 1996. Soxhlet Extraction (Revision 3, December 1996).
- 29.- EPA 3545. 1996. Pressurized fluid extraction (PFE). (Revisión 0, Diciembre de 1996).
- 30.- EPA 3535. 1996. Solid-phase extraction (SPE). (Revisión 0, Diciembre de 1996).
- 31.-EPA 3550 B. 1996. Ultrasonic extraction). (Revisión 2, Diciembre de 1996).
- 32.- EPA 3560. 1996. Supercritical fluid extraction of total recoverable Petroleum hydrocarbons). (Revisión 0, Diciembre de 1996).
- 33.- Erikson, M., Fäldt J., Dalhammar G., and Borg-Karlson A.K. 2001. Determination of hydrocarbons in old creosote contaminated soil using headspace solid phase microextraction and GC-MS. Chemosphere 44:1641-1648.
- 34.- Fernández-Linares., L. 2000. Tecnologías de biorremediación en suelos. p. 700-705. In: Quintero-Lizaola R. Reyna-Trujillo T, Corlay-Chee L., Ibáñez-Huerta A. y García-Calderón (eds.). La edafología y sus perspectivas al siglo XXI. Tomo II Colegio de Postgraduados, Universidad Autónoma de México, Universidad Autónoma Chapingo, México 700-705 p.
- 35.- Flores, P. M.A., Torras O. S., Téllez G.R. 2004. Medidas de mitigación para uso de suelos contaminados por derrames de hidrocarburos en infraestructura de transporte terrestre. Publicación Técnica No 257 Sanfandila, Qro., México

- 36.- Flores K, Sánchez Ma-T., Pérez-Marín D. Guerrero J.E y Garrido-Varo A. 2009. Feasibility in NIRS instruments for predicting internal quality in intact tomato. *Journal of Food Engineering* 91 311–318.
- 37.- Forina, M., 1993. *Introduzione alla Chimica Analitica con elementi di Chimimetria*. ECIG (Edizioni Culturali Internazionali Genova), 1ª edizione.
- 38.- García-López E, Zavala-Cruz J, Palma-López D.J. 2006. Caracterización de las comunidades vegetales en un área afectada por derrames de hidrocarburos. *Terra Latinoamericana* 24: 17-26.
- 39.- González, M. I., González-Pérez C., Hernández Méndez J., and Álvarez-García N. 2003. Determination of fatty acids in the subcutaneous fase of Iberian breed swine by near infrared spectroscopy (NIRS) with a fibre-optic probe. *Meat Science* 65:713-719.
- 40.- Goodacre, R. and Kell D. B. 1996. Correction of mass spectral drift using artificial neural networks. *Analytical Chemistry* 68:271-280
- 41.- Guerrero, P. A. 2006. Estudio de la Contaminación en Suelos por Petróleo Crudo Mediante Generación de Espacio de Cabeza-Espectrometría de Masas. Tesis de Doctorado. Universidad de Salamanca. Salamanca España.
- 42.- He Y., Huanga M., García A., Hernández A. and Songa H 2007. Prediction of soil macronutrients content using near-infrared spectroscopy. *Computers and Electronics in Agriculture* 58: 144-153.
- 43.- INEGI. Instituto Nacional de Estadística Geográfica e Informática. 2000. Anuario Estadístico del Estado de Tabasco. México D.F.

- 44.- IMP. 2008. Acerca del Petróleo. Tipos de petróleo. Información de internet. <http://www.imp.mx/petroleo>
- 45.- IUSS, ISRIC y FAO (2007) Base Referencial Mundial del Recurso Suelo. Primera actualización 2007. Informes sobre Recursos Mundiales de Suelos No. 103. FAO, Roma. 127 pp.
- 46.- Keller R. Mermet J.M., M Otto and Widmer H.M. 1998 Analytical Chemistry, John Wiley & Sons, New York
- 47.- Kesley, J., Kottler, B., y Alexander M. 1997. Selective chemical extractants to predict bioavailability of soli-aged organic chemicals. Environ. Sci. Technol. 31: 214-217
- 48.- Kostecki, P.T., Calabrese,E. y Simmons K. 2001. Survey of States' 2000 Soils Cleanup Standards for petroleum contamination. Soil and sediment Contamination 10:117-196.
- 49.- Lambert, P., Fingas M., and Goldthorp M. 2001. An evaluation of field total petroleum hydrocarbon (TPH) systems. Journal of Hazard Materials 83:65-81.
- 50.- Li, X., Feng, Y., y Sawatsky, N. 1997. Importance of soil-water relations in assessing the endpoint of bioremediated soils. Plant Soil 192: 219-226
- 51.- Liu Y., Chen X, and Ouyang A .2008. Nondestructive determination of pear internal quality indices by visible and near-infrared spectrometry. LWT - Food Science and Technology 28 (10) 2318-2320.
- 52.- Luque-García J. L. and Luque de C M. D.. 2001. Water Soxhlet extraction assisted by focused microwaves: a clean approach. Analytical Chemistry 73:5903-5908.

- 53.- Macho, S. A. 2002. Espectroscopia de infrarrojo y calibración multivariante. Aplicación a la industria petroquímica. Tesis doctoral. Universidad ROVIRA I VIRGILI. Tarragona, España. pp.23-63
- 54.- Macho, S. and M. S. Larrechi. 2002. Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. *Trends in Analytical Chemistry* 21:799-806
- 55.- Marthens H. N. T. 1996. *Multivariate Calibration*. John Wiley & Sons, Chichester.68:3851-3858.
- 56.- Massart D. L., Vandeginste G. M., Buydens L. M. C., De Jong S., Lewi P. J., and Smeyers-Verbeke J.1997. *Handbook of Chemometrics and Qualimetrics: Part A*. Elsevier, Amsterdam.
- 57.- Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman, L. 1988. *Data Handling in Science and Technology, volume 2. Chemometrics: a textbook*. Elsevier Science Publishers, Amsterdam, The Netherlands.
- 58.- Miller, J. N.; Miller J.C., 2000. *Statistic and Chemometrics for Analytical Chemistry*. Pearson Education, Harlow. pp 271
- 59.- Morra, M.J., Hall, M.H., Freeborn, L.L., 1991. Carbon and Nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy. *Soil Sci. Soc. Am. J.* 55:288-291.
- 60.- Nava, D. J. y Tirado, S. D. 1980. Efectos residuales en suelos por derrames de petróleo crudo. *In: Memoria XIII Congreso Nacional de la Ciencia del Suelo. Tomo I. Sociedad Mexicana de la Ciencia del Suelo. Toluca, edo de México.* pp 127-140.

- 61.- NOM-021-SEMARNAT-2000, Norma Oficial Mexicana que establece las especificaciones de fertilidad, salinidad y clasificación de suelos. Estudio, muestreo y análisis. Emitida en el Diario Oficial de la Federación el 23 de Abril de 2003.
- 62.- NOM-138-SEMARNAT/SS-2003. Norma oficial Mexicana Límites Máximos permisibles de hidrocarburos en suelos y sus especificaciones para su caracterización y remediación. Diario Oficial de la Federación 29 de Marzo de 2005.
- 63.- Palma-López D., J. Obrador, J. Zavala, E. García, C. Ortiz, W. Camacho, r. Meseguer, A. Sol, J. Juárez, J. Jasso, A. Guerrero, R. Ramos y A. Triano. 1999. Diagnostico de los recursos naturales, niveles de contaminación y alternativas para el desarrollo del área de influencia de los campos petroleros Cinco Presidentes y la Venta Norte. Colegio de Postgraduados H. Cárdenas, Tabasco, México.
- 64.- Palma-López D.J., Cisneros D.J., Moreno C.E., Rincón-Ramírez J.A. 2007. Suelos de Tabasco: Su uso y manejo sustentable. Colegio de Postgraduados-ISPROTAB-FUPROTAB. Villahermosa, Tabasco, México.
- 65.- Pérès, C., Viallon C., and Berdagué J. L. 2002. Curie point pyrolysis-mass spectrometry applied to rapid characterization of cheeses. Journal of Analytical and Applied Pyrolysis 65:161-171.
- 66.- Perrotta, G. and Xeferis P. 1996. Small synthetic aperture radar satellite constellations for tracking oil spills. Spill Science & Technology Bulletin 3:73-82.
- 67.- Physical/chemical methods. 1986. Test Methods for Evaluating Solid Waste. USEPA publication (SW-846 on line). Third edition. Environmental Protection Agency-United State of America.

- 68.- Pirouette: Comprehensive Chemometrics modeling Software, Infometrix, Ver. 3.0, Woodinville, WA, 2007.
- 69.- PROFEPA, 2000. Disposición y procedimientos para la caracterización y restauración de suelos contaminados. Memoria GDT (1998-2000) sobre restauración de suelos contaminados. México, D. F. 147 p.
- 70.- PROFEPA. 2003 Reporte de Emergencias ambientales en México. Procuraduría Federal de Protección al Ambiente. (PROFEPA). México. D.F.
- 71.- PROFEPA. 2009. Emergencias ambientales segundo trimestre de 2005. <http://www.profepa.gob.mx/PROFEPA/Emergencias+Ambientales/AnalisisNacionaldeEmergenciasAmbientales/Tabasco.htm>. Consultada el 20/11/2009.
- 72.- Radovic, B. S., Goodacre R., and Anklam E. 2001. Contributions of pyrolysis-mass spectrometry (Py-MS) to authenticity testing of honey. *Journal of Analytical Applied Pyrolysis* 60:79-87.
- 73.- Ramis R G., García Álvarez-Coque M. C. 2001. *Quimiometría. Síntesis*, Madrid
- 74.- Reeves G. 2005. *Understanding and monitoring hydrocarbons in water*. Oakville, Ontario, Canada, Arjay Engineering LTD.
- 75.- Rivera-Cruz M. del C., Ferrera-Cerrato R., Volke-Haller V., Fernández-Linares L and. Rodríguez-Vázquez R. 2002. Poblaciones microbianas en perfiles de suelos afectados por hidrocarburos del petróleo en el estado de Tabasco, México. *Agrociencia* 36:149-160.

- 76.- Schmidt, D E. 1999. Historical overview of oil spills from all sources (1960-1998). *In*: International Oil Spill Conference. 8-11 March. Seattle, USA.
- 77.- Schmidt, D. E. 2001. Analysis of oil spill trends in the United States and worldwide. *In*: International Oil Spill Conference. 26-29 March. Tampa, Florida, USA
- 78.- Skoog, D. A., West, D. M. 1989. Análisis instrumental. Segunda edición. Ed. McGraw Hill. México. Traducción al español, Mario Calcagno.
- 79.- Szolar, O. H. J., Rost H., Braun R., and Loibner A. P. 2002. Analysis of polycyclic aromatic hydrocarbons in soil. Minimizing sample pre-treatment using automated Soxhlet with ethyl acetate as extraction solvent. *Analytical Chemistry* 74:2379-2385.
- 80.- Thurston, T. J., Brereton R. G., Foord D. J., and Escott R. E. A. 2004. Principal components plots for exploratory investigation of reactions using ultraviolet-visible spectroscopy: application to the formation of benzophenone phenylhydrazone. *Talanta* 63:757-769
- 81.- Valcárcel, M. and Cárdenas S Ma. 2000. Automatización y miniaturización en química analítica. Springer – Verlag Ibérica S. A. Barcelona, España
- 82.- Velásquez E, Lavelle P., Barrios E., Joffre R., Reversat F. 2005. Evaluating soil quality in tropical agroecosystems of Colombia using NIRS. *Soil Biology & Biochemistry* 37.
- 83.- Wang, Z., Dean T., and Kowalski B. R. 1995. Additive background correction in multivariate instrument standardization. *Analytical Chemistry* 67:2379-2385.

- 84.- Wang, Z. and Fingas M.F. 2003. Development of oil hydrocarbon fingerprintand identification techniques. *Marine Pollution Bulletin* 47:423-452.
- 85.- Wold, S. 1976 Pattern Recognition by means of disjoint principal components models. *Pattern. Recog.* 8: 127-139.
- 86.- Zavala-Cruz, J, Gavi-Reyes, F, Adams-Schroeder, R.H, Ferrera-Cerrato, R, Palma-López, D J, Vaquera-Huerta, H, Domínguez-Ezquivel, J.M. 2005. Derrames de petróleo en suelos y adaptación de pastos tropicales en el Activo Cinco Presidentes, Tabasco, México. *Terra Latinoamericana* 23: 293-302.
- 87.- Zhang, L, Small C. W., and Arnold M. A. 2002. Calibration standardization algorithm for partial least-squares regression: application to the determination of physiological levels of glucose by near-infrared spectroscopy. *Analytical Chemistry* 74:4097-4108.

Cuadro 2. Muestras colectadas en la Cuenca baja del Río Tonalá

No. muestra	CLAVE DE CAMPO	Descripción	OBSERVACIONES
1	HS1-1	Histosol soditérico con manglar	Derrame de petróleo
2	HS1-2	Histosol soditérico con manglar	Derrame de petróleo
3	HS1-3	Histosol soditérico con manglar	Derrame de petróleo
4	HS1-4	Histosol soditérico con manglar	Derrame de petróleo
5	HS2-1	Histosol soditérico con manglar	Orilla de campo petrolero
6	HS2-2	Histosol soditérico con manglar	Orilla de campo petrolero
7	HS2-3	Histosol soditérico con manglar	Orilla de campo petrolero
8	HS2-4	Histosol soditérico con manglar	Orilla de campo petrolero
9	HS3-1	Histosol soditérico con manglar	Salida de la cuenca (desembocadura Río Tonalá)
10	HS3-2	Histosol soditérico con manglar	Salida de la cuenca (desembocadura Río Tonalá)
11	HS3-3	Histosol soditérico con manglar	Salida de la cuenca (desembocadura Río Tonalá)
12	HS3-4	Histosol soditérico con manglar	Salida de la cuenca (desembocadura Río Tonalá)
13	HSf1-1	Histosol fibrico con vegetación hidrófita	Derrame de petróleo
14	HSf1-2	Histosol fibrico con vegetación hidrófita	Derrame de petróleo
15	HSf1-3	Histosol fibrico con vegetación hidrófita	Derrame de petróleo
16	HSf1-4	Histosol fibrico con vegetación hidrófita	Derrame de petróleo
17	HSf2-1	Histosol fibrico con vegetación hidrófita	Orilla de campo petrolero
18	HSf2-2	Histosol fibrico con vegetación hidrófita	Orilla de campo petrolero
19	HSf2-3	Histosol fibrico con vegetación hidrófita	Orilla de campo petrolero
20	HSf2-4	Histosol fibrico con vegetación hidrófita	Orilla de campo petrolero
21	HSf3-1	Histosol fibrico con vegetación hidrófita	Salida de la cuenca (desembocadura Río Tonalá)
22	HSf3-2	Histosol fibrico con vegetación hidrófita	Salida de la cuenca (desembocadura Río Tonalá)
23	HSf3-3	Histosol fibrico con vegetación hidrófita	Salida de la cuenca (desembocadura Río Tonalá)
24	HSf3-4	Histosol fibrico con vegetación hidrófita	Salida de la cuenca (desembocadura Río Tonalá)
25	<u>HSf4-1</u>	Histosol fibrico con vegetación hidrófita	Testigo
26	<u>HSf4-2</u>	Histosol fibrico con vegetación hidrófita	Testigo
27	<u>HSf4-3</u>	Histosol fibrico con vegetación hidrófita	Testigo
28	<u>HSf4-4</u>	Histosol fibrico con vegetación hidrófita	Testigo
29	SC1-1	Solonchac sodi-gleyico con manglar	Derrame de petróleo
30	SC1-2	Solonchac sodi-gleyico con manglar	Derrame de petróleo
31	SC1-3	Solonchac sodi-gleyico con manglar	Derrame de petróleo

Cuadro 2. Continuación

32	SC1-4	Solonchac sodi-gleyico con manglar	Derrame de petróleo
33	SC2-1	Solonchac sodi-gleyico con manglar	Orilla de campo petrolero
34	SC2-2	Solonchac sodi-gleyico con manglar	Orilla de campo petrolero
35	SC2-3	Solonchac sodi-gleyico con manglar	Orilla de campo petrolero
36	SC2-4	Solonchac sodi-gleyico con manglar	Orilla de campo petrolero
37	SC3-1	Solonchac sodi-gleyico con manglar	Salida de la cuenca (desembocadura Río Tonalá)
38	SC3-2	Solonchac sodi-gleyico con manglar	Salida de la cuenca (desembocadura Río Tonalá)
39	SC3-3	Solonchac sodi-gleyico con manglar	Salida de la cuenca (desembocadura Río Tonalá)
40	SC3-4	Solonchac sodi-gleyico con manglar	Salida de la cuenca (desembocadura Río Tonalá)
41	<u>SC4-1</u>	Solonchac sodi-gleyico con manglar	Testigo
42	<u>SC4-2</u>	Solonchac sodi-gleyico con manglar	Testigo
43	<u>SC4-3</u>	Solonchac sodi-gleyico con manglar	Testigo
44	<u>SC4-4</u>	Solonchac sodi-gleyico con manglar	Testigo
45	GL1-1	Gleysol mólico con pastizal	Derrame de petróleo*
46	GL1-2	Gleysol mólico con pastizal	Derrame de petróleo
47	GL1-3	Gleysol mólico con pastizal	Derrame de petróleo
48	GL1-4	Gleysol mólico con pastizal	Derrame de petróleo
49	GL2-1	Gleysol mólico con pastizal	Orilla de campo petrolero
50	GL2-2	Gleysol mólico con pastizal	Orilla de campo petrolero
51	GL2-3	Gleysol mólico con pastizal	Orilla de campo petrolero
52	GL2-4	Gleysol mólico con pastizal	Orilla de campo petrolero
53	GL3-1	Gleysol mólico con pastizal	Salida de la cuenca (desembocadura Río Tonalá)
54	GL3-2	Gleysol mólico con pastizal	Salida de la cuenca (desembocadura Río Tonalá)
55	GL3-3	Gleysol mólico con pastizal	Salida de la cuenca (desembocadura Río Tonalá)
56	GL3-4	Gleysol mólico con pastizal	Salida de la cuenca (desembocadura Río Tonalá)
57	<u>GL4-1</u>	Gleysol mólico con pastizal	Testigo
58	<u>GL4-2</u>	Gleysol mólico con pastizal	Testigo
59	<u>GL4-3</u>	Gleysol mólico con pastizal	Testigo
60	<u>GL4-4</u>	Gleysol mólico con pastizal	Testigo
61	RG1-1	Arenosol háplico con pastizal y cocotero	Derrame de petróleo
62	RG1-2	Arenosol háplico con pastizal y cocotero	Derrame de petróleo
63	GL1-3	Arenosol háplico con pastizal y cocotero	Derrame de petróleo
64	RG1-4	Arenosol háplico con pastizal y cocotero	Derrame de petróleo
65	RG2-1	Arenosol háplico con pastizal y cocotero	Orilla de campo petrolero
66	RG2-2	Arenosol háplico con pastizal y cocotero	Orilla de campo petrolero

Cuadro 2. continuación

67	RG2-3	Arenosol háplico con pastizal y cocotero	Orilla de campo petrolero
68	RG2-4	Arenosol háplico con pastizal y cocotero	Orilla de campo petrolero
69	RG3-1	Arenosol háplico con pastizal y cocotero	Salida de la cuenca
70	RG3-2	Arenosol háplico con pastizal y cocotero	Salida de la cuenca
71	RG3-3	Arenosol háplico con pastizal y cocotero	Salida de la cuenca
72	RG3-4	Arenosol háplico con pastizal y cocotero	Salida de la cuenca
73	<u>RG4-1</u>	Arenosol háplico con pastizal y cocotero	Testigo
74	<u>RG4-2</u>	Arenosol háplico con pastizal y cocotero	Testigo
75	<u>RG4-3</u>	Arenosol háplico con pastizal y cocotero	Testigo
76	<u>RG4-4</u>	Arenosol háplico con pastizal y cocotero	Testigo

Cuadro 3. Propiedades físicas y químicas de los suelos analizados.

Solicitud: 002
Recepción de muestras: 03/01/08
Entrega de resultados : 03/02/09

Solicitante: DR. JOEL ZAYALA CRUZ
Asunto: Informe de Resultados de Análisis de Suelos

No.	Identificación	pH (H ₂ O)	CE dS m ⁻¹	MO %	N	K	Ca	Mg	Na	CIC	Arcilla	Limo	Arena	Clasificación
2	HS1-1	4.99	7.21	87.6	1.78	1.42	29.0	NSD	38.1	107.8	NA	NA	NA	NA
3	HS1-2	5.02	11.1	87.5	1.78	1.72	27.8	NSD	NSD	98.8	NA	NA	NA	NA
4	HS1-3	4.95	10.2	86.3	1.71	1.46	23.5	47.9	46.3	97.8	NA	NA	NA	NA
5	HS1-4	4.86	9.20	86.9	1.85	1.52	25.9	51.6	45.1	95.8	NA	NA	NA	NA
6	HS2-1	4.87	17.7	86.7	1.57	3.07	25.1	NSD	NSD	97.8	NA	NA	NA	NA
7	HS2-2	4.90	21.5	82.5	1.47	3.49	26.0	NSD	NSD	99.8	NA	NA	NA	NA
8	HS2-3	4.93	20.5	83.0	1.43	3.49	21.2	NSD	NSD	99.8	NA	NA	NA	NA
9	HS2-4	5.01	20.9	83.7	1.43	3.06	26.5	NSD	NSD	90.8	NA	NA	NA	NA
10	HS3-1	4.50	14.4	46.5	0.52	2.06	12.0	29.8	NSD	56.9	NA	NA	NA	NA
11	HS3-2	4.48	8.79	31.0	0.31	0.85	5.58	12.2	35.6	26.0	NA	NA	NA	NA
12	HS3-3	4.67	7.54	34.4	0.45	1.60	8.27	22.0	NSD	50.9	NA	NA	NA	NA
13	HS3-4	4.63	5.48	28.3	0.28	0.75	5.09	10.5	31.7	22.0	NA	NA	NA	NA
14	HSf1-1	5.32	0.82	29.3	0.38	0.15	16.6	5.10	3.22	28.0	25	19	56	Migajón Arcillo-Arenoso
15	HSf1-2	4.86	0.92	43.8	0.63	0.08	10.5	3.80	1.04	36.9				
16	HSf1-3	6.48	0.42	17.0	0.24	0.09	15.7	4.62	1.05	18.0	37	29	34	Migajón Arcilloso

17	HSf1-4	5.74	0.42	20.4	0.28	0.11	13.0	4.28	1.39	29.9	41	29	30	Arcilla
18	HSf2-1	5.25	0.20	29.5	0.59	0.13	16.3	6.25	0.78	46.9	29	25	46	Migajón Arcillo-Arenoso
19	HSf2-2	4.65	0.90	68.7	1.61	0.29	29.8	11.5	2.26	80.9	ND	ND	ND	ND
20	HSf2-3	4.50	0.97	88.7	2.24	0.45	27.1	8.72	2.00	89.8	ND	ND	ND	ND
21	HSf2-4	4.33	0.98	90.3	2.34	0.56	28.5	9.38	1.91	81.9	ND	ND	ND	ND
22	HSf3-1	4.58	1.78	70.1	1.57	0.38	10.5	18.1	18.7	75.9	ND	ND	ND	ND
23	HSf3-2	4.46	2.31	87.5	2.10	0.42	13.6	21.7	19.2	94.8	ND	ND	ND	ND
24	HSf3-3	4.53	1.71	88.0	2.10	0.46	12.0	19.1	15.6	30.9	ND	ND	ND	ND
25	HSf3-4	4.63	1.72	86.9	2.10	0.49	16.8	25.2	15.8	85.9	ND	ND	ND	ND
26	HSf4-1	4.47	0.69	84.0	2.24	0.52	22.9	10.0	1.48	86.9	ND	ND	ND	ND
27	HSf4-2	4.50	0.68	89.3	2.31	0.50	26.3	10.9	7.65	95.8	ND	ND	ND	ND
28	HSf4-3	4.47	0.74	89.5	2.31	0.46	28.6	11.2	1.37	94.8	ND	ND	ND	ND
29	HSf4-4	4.43	1.00	88.3	2.31	0.54	28.9	11.0	2.04	83.9	ND	ND	ND	ND
30	SC1-1	ND	ND	ND	ND	ND								
31	SC1-2	ND	ND	ND	ND	ND								
32	SC1-3	6.33	7.98	70.2	1.47	1.50	8.82	26.2	42.4	83.9	ND	ND	ND	ND
33	SC1-4	ND	ND	ND	ND	ND								
34	SC2-1	5.19	8.21	34.0	0.56	1.85	9.54	21.4	41.7	42.9	41	23	36	Arcilla
35	SC2-2	5.43	6.98	29.5	0.42	1.80	4.29	22.5	38.2	38.9	45	19	36	Arcilla
36	SC2-3	5.32	7.70	31.2	0.49	1.92	5.39	26.0	41.8	40.9	45	19	36	Arcilla
37	SC2-4	5.36	7.22	33.5	0.52	1.83	7.50	23.0	38.8	40.9	41	19	40	Arcilla
38	SC3-1	4.67	9.00	36.5	0.31	1.86	7.15	20.1	45.8	44.9	29	13	58	Migajón Arcillo-Arenoso

39	SC3-2	4.69	8.17	39.5	0.49	1.75	7.77	19.4	42.6	47.9	25	13	62	Migajón Arcillo-Arenoso
40	SC3-3	4.59	8.63	33.0	0.42	1.84	7.24	21.7	48.7	46.9	27	11	62	Migajón Arenoso
41	SC3-4	4.66	7.06	28.5	0.28	1.34	6.16	17.1	41.1	36.9	19	11	70	Migajón Arenoso
42	SC4-1	4.18	9.10	15.8	0.38	1.82	9.87	21.1	41.3	33.9	29	23	48	Migajón Arcillo-Arenoso
43	SC4-2	3.76	9.01	15.1	0.35	1.63	8.23	21.2	42.0	31.9	23	33	44	Franco
44	SC4-3	3.76	10.0	17.4	0.42	1.58	9.19	21.7	43.3	31.9	29	35	36	Migajón Arcilloso
45	SC4-4	4.11	8.07	13.1	0.35	1.40	7.12	16.1	41.9	27.0	27	35	38	Migajón Arcilloso
46	GL1-1	6.12	94.9	2.10	0.07	1.56	8.11	5.13	0.35	18.0	33	17	50	Migajón Arcillo-Arenoso
47	GL1-2	5.98	0.14	2.56	0.10	0.13	7.61	4.29	0.38	15.0	31	21	48	Migajón Arcillo-Arenoso
48	GL1-3	6.18	0.12	2.63	0.14	0.18	11.6	4.28	0.97	23.0	45	19	36	Arcilla
49	GL1-4	6.43	0.10	3.09	0.10	0.19	12.6	5.43	0.81	23.0	45	25	30	Arcilla
50	GL2-1	5.49	0.08	5.19	0.21	0.12	8.90	6.74	1.02	28.0	55	27	18	Arcilla
51	GL2-2	4.89	0.07	6.04	0.28	0.15	9.97	7.40	0.97	29.0	55	33	12	Arcilla
52	GL2-3	5.51	0.07	4.93	0.28	0.13	9.93	6.09	0.96	27.0	53	31	16	Arcilla
53	GL2-4	5.48	0.06	4.93	0.24	0.13	9.22	5.93	1.01	28.0	55	31	14	Arcilla
54	GL3-1	5.72	0.35	4.07	0.17	0.23	6.24	8.72	3.79	24.0	49	37	14	Arcilla
55	GL3-2	5.84	0.29	3.68	0.17	0.20	5.81	9.70	3.40	23.0	45	43	12	Arcilla Limosa
56	GL3-3	5.61	0.29	4.66	0.24	0.22	5.50	10.7	3.32	24.0	48	43	9	Arcilla Limosa
57	GL3-4	5.61	0.33	4.34	0.21	0.23	5.74	11.2	3.34	26.0	50	41	9	Arcilla Limosa
58	GL4-1	5.86	0.07	4.60	0.24	0.20	13.1	8.06	0.77	30.9	62	35	3	Arcilla
59	GL4-2	5.68	0.07	6.11	0.28	0.19	11.9	7.24	0.39	29.0	56	39	5	Arcilla
60	GL4-3	5.69	0.07	5.85	0.28	0.18	11.8	7.57	0.38	29.9	58	37	5	Arcilla

61	GL4-4	5.69	0.06	5.45	0.28	0.17	11.7	0.71	0.77	28.0	56	35	9	Arcilla
62	RG1-1	5.54	0.04	2.30	0.10	0.05	0.95	0.49	0.05	3.99	10	7	83	Arena Migajosa
63	RG1-2	5.76	0.03	2.10	0.07	0.04	0.82	0.35	0.20	2.99	10	5	85	Arena Migajosa
64	GL1-3	5.96	0.04	2.04	0.10	0.05	1.43	0.39	0.34	3.99	10	5	85	Arena Migajosa
65	RG1-4	5.30	0.05	6.04	0.10	0.04	0.98	0.39	0.06	4.99	10	5	85	Arena Migajosa
	RG2-1	6.11	0.04	2.36	0.14	0.13	1.20	0.90	0.08	4.99	10	3	87	Arena Migajosa
67	RG2-2	5.93	0.04	2.43	0.10	0.17	1.35	0.90	0.07	5.99	10	5	85	Arena Migajosa
68	RG2-3	5.72	0.03	2.69	0.14	0.06	0.89	0.41	0.05	4.99	10	5	85	Arena Migajosa
69	RG2-4	5.82	0.03	2.89	0.14	0.07	0.74	0.44	0.10	4.99	1	4	95	Arena
70	RG3-1	6.29	0.03	1.31	0.10	0.13	1.24	0.99	0.08	2.99	8	3	89	Arena
71	RG3-2	6.05	0.03	1.38	0.10	0.11	1.14	0.69	0.05	2.99	8	3	89	Arena
72	RG3-3	6.10	0.03	1.31	0.07	0.12	1.28	0.90	0.06	4.99	8	3	89	Arena
73	RG3-4	5.99	0.03	1.51	0.10	0.11	1.29	0.81	0.21	3.99	8	3	89	Arena
74	RG4-1	6.56	0.04	0.99	0.07	0.08	0.94	1.32	0.12	2.99	8	3	89	Arena
75	RG4-2	6.78	0.04	0.79	0.07	0.10	0.60	1.38	0.09	2.00	8	1	91	Arena
76	RG4-3	6.97	0.04	0.99	0.07	0.09	0.73	1.37	0.13	2.00	8	1	91	Arena
77	RG-4	6.84	0.04	1.18	0.10	0.14	1.02	1.77	0.14	2.00	8	1	91	Arena

ND= No determinada

Métodos: Los análisis se realizaron de acuerdo a la Norma Oficial Mexicana NOM-021-RECNAT-2000, que establece las especificaciones de fertilidad, salinidad y clasificación de suelos.

Cualquier aclaración favor de realizarla antes del 03 de Febrero de 2009

ccp.- Archivo

Cuadro 4. Puntuaciones de las componentes principales del espectro NIR del suelo Histosol.

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8
HS-derrame	-1.068802	0.179017	0.116743	0.010054	-0.018782	-0.041298	0.015521	0.010054
HS-derrame	-0.227119	1.445079	0.331348	0.007345	0.046739	0.032724	-0.02011	-0.004804
HS-derrame	-0.808905	0.414915	0.169765	0.044403	-0.044856	-0.02566	-0.001145	-0.006081
HS-derrame	-1.058745	0.255572	0.105962	-0.01337	-0.005529	-0.024487	0.008803	-0.008029
HS-orilla campo	-1.471129	0.268025	-0.078896	-0.003614	-0.017785	0.003991	0.019685	0.004952
HS-orilla campo	-1.674137	0.303253	-0.106929	0.026836	-0.044483	0.004821	-0.013457	0.005579
HS-orilla campo	-1.761968	0.232186	-0.109484	-0.018497	-0.039206	0.013436	-0.013217	0.001032
HS-orilla campo	-1.861118	0.192035	-0.106164	-0.035353	-0.040683	0.006337	-0.030002	-0.001762
HS-desembocadura	-1.112804	0.437867	-0.083844	0.021045	0.035374	0.009012	-0.012835	0.006447
HS-desembocadura	-0.556382	0.090305	-0.170016	0.023173	0.063581	-0.011874	0.01375	0.000178
HS-desembocadura	-0.592685	0.154088	-0.146233	0.022893	0.059577	-0.003933	0.018318	-0.00643
HS-desembocadura	-0.080203	0.349873	-0.135051	0.058832	0.079362	-0.002078	0.010273	-0.013192
HSf-derrame	-0.07341	-1.086099	0.117881	0.021279	0.047569	-0.017307	-0.014363	0.007972
HSf-derrame	0.096372	-0.688779	0.176133	-0.022044	0.029273	-0.025087	-0.012225	0.000909
HSf-derrame	-0.607227	-1.599324	0.064491	0.006932	0.039848	0.026965	-0.008649	-0.001938
HSf-derrame	-0.477087	-1.718483	0.005339	0.015132	-0.01627	0.013726	-0.004223	-0.003224
HSf-orilla campo	0.82352	-0.267707	0.069527	0.185361	-0.068885	0.022595	0.024535	-0.000433
HSf-orilla campo	0.767087	0.107045	0.025491	0.019808	-0.011678	0.013773	0.019578	0.009217
HSf-orilla campo	0.289307	0.250356	0.109488	-0.078628	0.010678	0.011916	0.022933	0.005369
HSf-orilla campo	0.091496	0.025728	0.04037	-0.140798	0.001263	0.025204	0.026438	0.003798
HSf-desembocadura	1.604165	0.028062	-0.087166	0.019454	-0.004635	-0.001327	-0.007985	0.00535
HSf-desembocadura	1.470788	0.171468	-0.072432	-0.021434	0.013564	-0.006298	-0.003175	0.014185
HSf-desembocadura	1.913965	0.449897	-0.022282	0.020198	0.007262	-0.000452	-0.01622	-0.00069
HSf-desembocadura	2.056635	0.426857	-0.052226	0.028582	-0.001548	-0.002493	-0.022175	0.002534
HSf-testigo	0.681506	-0.145969	0.029504	-0.064039	-0.034686	0.000184	0.011037	-0.013569
HSf-testigo	0.951328	-0.275308	-0.077302	-0.088126	-0.029142	-0.007751	-0.006368	-0.002163
HSf-testigo	1.221644	-0.146395	-0.068262	-0.040254	-0.037022	-0.008887	-0.003844	-0.013558
HSf-testigo	1.463908	0.146437	-0.045755	-0.00517	-0.0189	-0.005753	-0.000879	-0.001704