

**CONSTRUCCIÓN DE UN MODELO PREDICTIVO DE LA FRACCIÓN S.A.R.A.
DE CRUDOS COLOMBIANOS A PARTIR DE ESPECTROSCOPIA (LIF) Y
MÉTODOS ESTADÍSTICOS (QUIMIOMETRÍA)**

WILMAR ANTONIO CONTRERAS TOLOZA

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
BUCARAMANGA
2011**

**CONSTRUCCIÓN DE UN MODELO PREDICTIVO DE LA FRACCIÓN S.A.R.A.
DE CRUDOS COLOMBIANOS A PARTIR DE ESPECTROSCOPIA (LIF) Y
MÉTODOS ESTADÍSTICOS (QUIMIOMETRÍA)**

WILMAR ANTONIO CONTRERAS TOLOZA

**Trabajo de grado para optar al título de
QUIMICO**

**Director:
ENRIQUE MEJÍA OSPINO
Químico, Ph. D.**

**Co-director:
JORGE ARMANDO ORREGO RUÍZ
Químico, M. Sc.**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
BUCARA MANGA
2011**

*"A mi queridísima Madre, a mi padre, a mi nona, y a mis
hermanas"*

AGRADECIMIENTOS

Agradezco a todas aquellas personas y entidades que han contribuido en mi formación personal y profesional, y especialmente a:

Mi madre, Flor Angela Toloza Villamizar, por sus sacrificios que jamás olvidaré, su preocupación, por creer en mí, y por su voz de aliento y amor en los momentos duros.

Mi nona, Florinda Villamizar Villabona, por ser la abuela mas dulce, por su colaboración, su preocupación, por ser otra madre para mí.

Mi padre, Luis María Contreras, por la alegría que expresó cada vez que superé una meta.

Mis hermanas, Ruth Sary e Irayda del Pilar, por su compañía.

A Yaqueline Caballero Plata, por su apoyo y preocupación.

A mis tías, Elida y Luz Marina Toloza Villamizar, por hacer que en principio todo esto fuera posible.

El doctor Enrique Mejía Ospino, por darme la oportunidad de trabajar a su lado, por sus valiosas enseñanzas, y por ser un ejemplo como persona y científico.

Al Profesor Rafael Cabanzo Hernández, por sus consejos, por su disposición a colaborar, por sus charlas y buen humor.

A mis compañeros de Laboratorio, Sara Parra, Johanna Gómez, Sonia Villamizar, Reggy Brieva, y Leonardo Muñoz, por su amistad, por las largas discusiones acompañadas de un buen café.

A Jorge Armando Orrego Ruíz, por su interés en mi trabajo, por los consejos, por su amistad y por sus críticas.

A la Universidad Industrial de Santander, especialmente a las escuelas de Química donde me formé, y a la escuela de Física por permitirme realizar mi trabajo de grado en el Laboratorio de Espectroscopia Atómica y Molecular (LEAM).

Al Instituto Colombiano del Petróleo (ICP), por la experiencia de haber podido trabajar con tan importante entidad.

CONTENIDO

	Pág
INTRODUCCIÓN	14
1. CONSIDERACIONES TEÓRICAS	16
1.1. EL PETRÓLEO	16
1.2. EL FENOMENO DE LA FLUORESCENCIA	18
1.3. QUIMIOMETRÍA	28
1.3.1. MÉTODOS PARA EL PRETRATAMIENTO ESPECTRAL.....	29
1.3.2. MÉTODOS PARA EL PRETRATAMIENTO DE DATOS.....	30
1.3.3. MÉTODOS PARA LA REDUCCIÓN DIMENSIONAL	31
1.3.4. MÉTODOS PARA EL RECONOCIMIENTO DE PATRONES	34
1.3.5. MÉTODOS PARA LA CONSTRUCCIÓN DE MODELOS PREDICTIVOS...36	
1.3.6. MÉTODOS PARA EVALUAR LA CAPACIDAD PREDICTIVA.....	40
2. PARTE EXPERIMENTAL	42
2.1. MONTAJE EXPERIMENTAL	42
2.1.1. COMPONENTES DEL MONTAJE EXPERIMENTAL	43
2.2. ADQUISICIÓN DE DATOS	46
2.3. PROCESAMIENTO DE DATOS	48
3. RESULTADOS Y ANÁLISIS	51
3.1. ANÁLISIS DE FACTORES PRINCIPALES (PFA)	51
3.2. ANÁLISIS DISCRIMINANTE LINEAL (LDA)	52
3.3. CONSTRUCCIÓN DEL MODELO PREDICTIVO (MLR).....	52
3.3.1. Elección de variables principales usando espectros auto-escalados.....	59
3.3.2. Elección de variables principales en espectros filtrados por FT.....	59

3.3.3. Resultados de la construcción del modelo predictivo	63
3.3.4. Otros resultados.....	66
4. CONCLUSIONES	68
5. RECOMENDACIONES.....	71
6. BIBLIOGRAFÍA	73
7. ANEXOS.....	80

LISTA DE FIGURAS

	Pág.
Figura 1. Esquema determinación de S.A.R.A.	17
Figura 2. Diagrama de energía parcial para un sistema fotoluminiscente.	19
Figura 3. Esquema de funcionamiento del PFA.....	33
Figura 4. Esquema de funcionamiento de LDA.....	36
Figura 5. Esquema fundamental de la quimiometría.....	37
Figura 6. Esquema experimental de LIF front-face.	42
Figura 7. Consola del Láser de nitrógeno PL 2300.....	43
Figura 8. Lente cilíndrico.....	44
Figura 9. Fibra óptica.....	44
Figura 10. Espectrómetro (USB4000 Ocean Optics).....	45
Figura 11. Celda de 1 cm.....	45
Figura 12. Espectro LIF del crudo C1.....	46
Figura 13. Espectros encontrados en la literatura.....	47
Figura 14. Espectros LIF de los 20 crudos	47
Figura 16. Corrección de línea base.	49
Figura 17. Espectros normalizados a longitud unidad.....	50
Figura 18. Resultados PFA. PF1 vs. PF2.....	51
Figura 19. Regresiones locales para saturados.....	54
Figura 20. Regresiones locales para aromáticos.....	55
Figura 21. Regresiones locales para resinas.....	56
Figura 22. Regresiones locales para asfaltenos.....	57
Figura 23. Espectros LIF de las fracciones S.A.R.A. de un crudo Venezolano.....	58

Figura 24. Espectros LIF auto-escalados	59
Figura 25. Espectros LIF sin filtrado por FT	60
Figura 26. Espectros LIF filtrados por FT.....	61
Figura 27. Elección de las variables principales para el crudo C3.....	62
Figura 28. Elección de las variables principales para el crudo C16	62
Figura 29. Elección de las variables principales para el crudo C12	63
Figura 30. Validación cruzada para variables obtenidas por FT-derivación en saturados.	64
Figura 31. Validación cruzada para variables obtenidas por FT-derivación en aromáticos.	64
Figura 32. Validación cruzada para variables obtenidas por FT-derivación para resinas.	65
Figura 33. Validación cruzada para variables obtenidas por FT-derivación en asfaltenos.	65
Figura 34. Reordenamiento de datos PFA.....	67

LISTA DE TABLAS

	Pág.
Tabla 1. Fracción S.A.R.A. para los 20 crudos (valores referencia)	53
Tabla 2. Errores para el modelo MLR.....	66

RESUMEN

TÍTULO

CONSTRUCCIÓN DE UN MODELO PREDICTIVO DE LA FRACCIÓN S.A.R.A. DE CRUDOS COLOMBIANOS A PARTIR DE ESPECTROSCOPIA (LIF) Y MÉTODOS ESTADÍSTICOS (QUIMIOMETRÍA).*

AUTOR

WILMAR ANTONIO CONTRERAS TOLOZA. **

PALABRAS CLAVES

LIF, CARA-FRONTAL, LDA, PCA, S.A.R.A.

DESCRIPCIÓN

La predicción de la fracción S.A.R.A. (Saturados, Aromáticos, Resinas, y Asfaltenos) basado en los espectros de fluorescencia inducida por laser (LIF) en modo de cara-frontal de veinte muestras de petróleo crudo colombiano y regresión lineal múltiple fue llevada a cabo. Para excitar la muestra, se usó la excitación de 337 nm de un laser de nitrógeno, la radiación de emisión fue colectada con una fibra óptica hasta un espectrómetro Ocean Optics modelo USB4000. La determinación experimental del análisis S.A.R.A. fue provista por el ICP (Instituto Colombiano del Petróleo) siguiendo normas ASTM. Este método tiene como principal ventaja omitir el uso de solventes, algo muy deseado en química analítica porque evitan los errores asociados a la preparación de la muestra. La matriz de datos espectroscópicos se construyó después de elegir como variables principales los puntos críticos de los espectros usando su primera derivada, previa extracción del ruido usando transformada de Fourier. La derivación se calculó por diferencias divididas. Mediante al análisis discriminante lineal (LDA) se estableció una ecuación a partir de las puntuaciones en el PC1 y PC2 calculados por análisis de componentes principales (PCA). Esta expresión permitió distinguir el origen geográfico de las muestras. La validación del modelo se realizó mediante validación cruzada cuyos errores calculados como RMSEV fueron 2.8934, 3.2790, 2.6720, y 1.2523 para saturados, aromáticos, resinas, y asfaltenos, respectivamente. Este modelo muestra una buena capacidad predictiva de las fracciones saturada, aromática, y resinas. Sin embargo, falla la predicción de asfaltenos.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Química. Director: Enrique Mejía Ospino. Codirector: Jorge Armando Orrego.

SUMMARY

TITLE

OBTENTION OF A PREDICTIVE MODEL OF THE S.A.R.A. FRACTION OF COLOMBIAN CRUDE OILS BY LIF SPECTROSCOPY AND STATISTIC METHODS (CHEMOMETRY).*

AUTHOR

WILMAR ANTONIO CONTRERAS TOLOZA.**

KEY WORDS

LIF, FRONT-FACE, LDA, PCA, S.A.R.A.

DESCRIPTION

The prediction of S.A.R.A. fraction (Saturates, Aromatics, Resins, and Asphaltene Content) based on spectra of laser induced fluorescence (LIF) on front-face mode of twenty raw Colombian crude oil and multiple linear regressions was performed. A nitrogen laser with light of 337 nm as excitation's source it was used. The radiation emitted was collected through an optical fiber in a spectrometer Ocean Optics model USB4000. The experimental determination of the S.A.R.A. analysis was provided by the ICP (in english, Colombian Petroleum Institute) following ASTM standards. This method has the main advantage omitting the use of solvents, something highly desirable in analytical chemistry because it is possible to avoid errors associated with sample preparation. The matrix of spectroscopic data was built after choosing as principal variables the critical points of the spectra using its first derived, previous subtraction of the noise using transformed of Fourier as filter. The derivation was calculated by divided differences. By linear discriminant analysis (LDA) one equation was established using the scores on PC1 and PC2 calculated by Principal Component Analysis (PCA). This expression allowed distinguishing the geographical origin of the samples. The validation of the model was done through cross validation whose errors in terms on RMSEV were 2.8934, 3.2790, 2.6720, and 1.2523, for saturates, aromatics, resins, and asphaltenes, respectively. This model shows a good predictive capacity for saturated, aromatic, and resins fractions; however, the prediction of asphaltenes fails.

* Work Degree

** Sciences Faculty. Chemistry School. Directress: Enrique Mejia Ospino. Codirectress Jorge Armando Orrego

INTRODUCCIÓN

Un recurso ampliamente estudiado en las últimas décadas es el petróleo, una mezcla cuya apariencia vira desde breas sólidas hasta líquidos claros, algunas veces pero no formalmente reconocida como una súper-mezcla y cuya composición no es definida; uno de los objetivos más deseados por los científicos e ingenieros del área es precisamente, determinar su composición, sin embargo, esta varía de un campo petrolero a otro, en un campo varía de un pozo a otro, y en un pozo varía de una ubicación a otra, haciendo imposible hablar de una composición definida para el petróleo; para superar esto, los estudiosos del petróleo han propuesto el parámetro S.A.R.A., el cual intenta dar cuenta de la composición del crudo en términos de S(Saturados) A(Aromáticos R(Resinas) A(Asfaltenos) y que actualmente se calcula con la ayuda de separaciones basadas en extracciones con solventes y cromatografía, involucrando costos elevados, largos tiempos de análisis, y la intervención de personal altamente calificado.

En el mundo industrial el petróleo genera el movimiento de miles de millones de dólares por día, esto demanda herramientas analíticas cada vez más sofisticadas con el objetivo de garantizar productos de excelente calidad en un mundo cada vez más acelerado y consumista. Por supuesto, la espectroscopia es la opción analítica que cada vez cobra más fuerza, no solo por ser rápida sino porque generalmente es asociada con un bajo costo, no requieren de personal altamente preparado, últimamente han sido fácilmente adaptada a tecnología portátil, y porque en las últimas décadas se ha incluido al LASER como fuente de excitación superando algunas limitaciones impuestas por el uso de lámparas. El petróleo puede considerarse como la muestra perfecta para ser estudiada por LIF (fluorescencia inducida por Laser), hecho que se confirma en muchas de las publicaciones que en revistas tales como *Fuels*, dedicadas exclusivamente al

estudio del petróleo, sus fracciones, y sus productos, muestran el uso de este fenómeno como medida para atender a diferentes inquietudes científicas a cerca de dicha mezcla.

No todas las moléculas tienen la capacidad de fluorescer, sin embargo, aquellas moléculas que la rodean (tanto las que fluorescen como las que no lo hacen) perturban la forma como la primera lo hace, poniendo en evidencia una manifestación colectiva de todas las moléculas que componen la mezcla, esto quiere decir que debería haber una correlación entre la composición y la medida de la fluorescencia; la cuestión es que hay muchas moléculas, lo que quiere decir que en el sentido de la composición real del petróleo, hay muchas variables, por supuesto, el fenómeno de fluorescencia involucra varios procesos y aplicado a una mezcla tan compleja como el petróleo, resulta en un problema en el que se trabajan muchas mas variables; por esto, en los últimos años se han reforzado los métodos espectroscópicos con métodos de análisis multivariable, fusión a la que desde aproximadamente los años 1970, se le denomina como quimiometría.

La ventaja de la quimiometría estriba en que propiedades difíciles de medir como S.A.R.A, pueden ser relacionadas a las medidas de LIF, permitiendo la posibilidad de calcular S.A.R.A. desde las medidas de fluorescencia inducida por laser, lo cual condensa la idea principal de esta monografía. Particularmente el presente trabajo muestra como se construye un modelo predictivo de la fracción S.A.R.A medida de acuerdo a las normas ASTM de algunos crudos Colombianos, a partir de sus espectros de fluorescencia inducida por laser.

1. CONSIDERACIONES TEÓRICAS

1.1. EL PETRÓLEO

También conocido como crudo, el petróleo es una mezcla cuya apariencia varía desde breas casi sólidas hasta líquidos aceitosos claros, como resultado de la metamorfosis de productos naturales sometidos a cambios químicos y físicos impartidos por las condiciones prevalecientes en un lugar particular. El crudo es un recurso al que en ocasiones se le ha considerado como una súper-mezcla y su composición está relacionada a la naturaleza y a la mezcla del material orgánico que lo generó, así como a la influencia de los procesos naturales como la biodegradación y al contacto con los diferentes componentes del suelo que rodeó el reservorio durante millones de años ¹, así la composición del petróleo obtenido de un pozo es variable y depende no solamente de la composición original del crudo in situ sino también en la manera de producción y la etapa alcanzada en la vida del pozo o reservorio.

Una consecuencia de lo dicho anteriormente es que el petróleo varía en composición de un campo a otro, de un pozo a otro en el mismo campo, y cada uno de un nivel a otro en el mismo pozo ². Esta variación puede ser en dos aspectos, el peso y el tipo de moléculas presentes en el petróleo. El petróleo podría ser descrito como una mezcla de moléculas orgánicas con una amplia distribución de tipos moleculares que caen dentro de una amplia distribución de pesos moleculares ³ y no se puede referir a él como una mezcla de composición definida.

Se sabe de la historia formal que la puesta en marcha de la extracción a escala industrial del crudo se remonta al 21 de agosto de 1859 en Pensilvania (U.S.A), y desde allí ha empezado a expandirse por el mundo la explotación de este recurso ⁴ así como de estudiarlo y entenderlo con el fin de aprovecharle mejor ⁵; quizá el parámetro que se desea saber con más ansia acerca del petróleo, sea su

composición, sin embargo, dada la complejidad de la misma y la necesidad de tener un parámetro que dé información acerca de ésta, los científicos e ingenieros propusieron un parámetro denominado fracción S.A.R.A. como una forma de expresar la composición de un crudo.

1.1.1. La Fracción S.A.R.A.

El parámetro S.A.R.A, o también denominado fracción S.A.R.A., se utiliza para expresar la composición gruesa de petróleos o fracciones pesadas del mismo y cuya sigla corresponde a S (Saturados), A (Aromáticos), R (Resinas), A (Asfaltenos).

La ASTM ofrece varios procedimientos para el análisis SARA: la norma ASTM D2007 y la ASTM D4124 mostrados en la figura 1.a y 1.b respectivamente, junto con la norma ASTM D2549 que usa la cromatografía líquida para la separación de las fracciones de saturados y aromáticos.

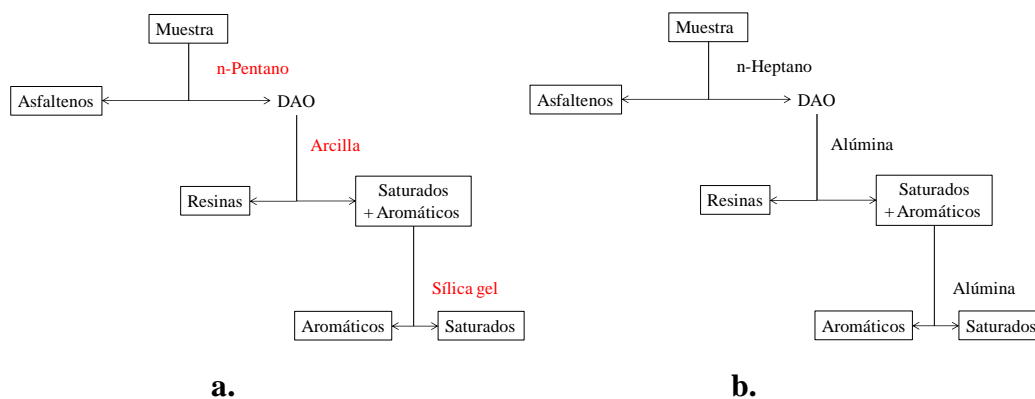


Figura 1. Esquema determinación de S.A.R.A.

Donde la parte del crudo que es soluble en pentano o heptano se le denomina maltenos. Esta última se puede fraccionar aun más por percolación a través de un material superficialmente activado con alúmina o arcilla. La fracción más

fuertemente adsorbida y que se desorbe con piridina o una mezcla tolueno-metanol se conoce como resinas. La fracción que eluye por la acción del tolueno o el benceno se define como aromáticos, mientras que la fracción que lo hace por la acción del heptano se conoce con el nombre de saturados.

1.2. EL FENÓMENO DE LA FLUORESCENCIA.

La luminiscencia es la emisión de luz producida por el decaimiento radiativo desde estados electrónicos excitados de una sustancia. La mayoría de las moléculas tienen sus espines apareados. Un estado electrónico en el cual todos los espines de los electrones están apareados se llama estado singulete. Cuando uno de los electrones de una molécula es excitado a un nivel de energía superior, se forma un estado singulete o triplete. En el estado singulete excitado, el espín del electrón promocionado continúa apareado con el electrón del estado fundamental; sin embargo, en el estado triplete los espines de los dos electrones se han desapareado. El estado triplete excitado es menos energético que el correspondiente estado singulete excitado, sin embargo, una transición singulete/triplete (o viceversa), es un suceso significativamente menos probable que la transición singulete/singulete, y como consecuencia de esto, el tiempo de vida medio de un estado triplete excitado puede oscilar desde 10^{-4} s a varios segundos, mientras que el tiempo de vida medio para un estado singulete excitado es de 10^{-8} a 10^{-5} s.

La foto-luminiscencia es el decaimiento radiativo desde estados electrónicos alcanzados después de un proceso de absorción de radiación electromagnética, hasta estados electrónicos de menor energía; esta se divide en dos categorías, la fluorescencia y la fosforescencia, dependiendo de la naturaleza del estado excitado. La fluorescencia es el proceso que se da entre estados de igual multiplicidad (proceso espín-permitido), así que la fosforescencia resulta de la transición entre estados de diferente multiplicidad.

Un diagrama de niveles de energía para una molécula foto-luminiscente característico, es ilustrado en la figura 2, donde la línea horizontal gruesa en la parte inferior representa la energía del estado fundamental de la molécula, que normalmente es un estado singulete y se designa S_0 . Las líneas gruesas superiores son los niveles de energía de los estados vibracionales fundamentales de los tres estados electrónicos excitados. S_1 y S_2 representan respectivamente los estados singulete excitados primero y segundo, y T_1 representa al primer estado triplete excitado. Las líneas horizontales delgadas que acompañan a las gruesas representan las energías de los estados vibracionales asociados con cada estado electrónico. Las líneas verticales delgadas que acompañan a las gruesas representan las energías de los estados vibracionales asociados con cada estado electrónico.

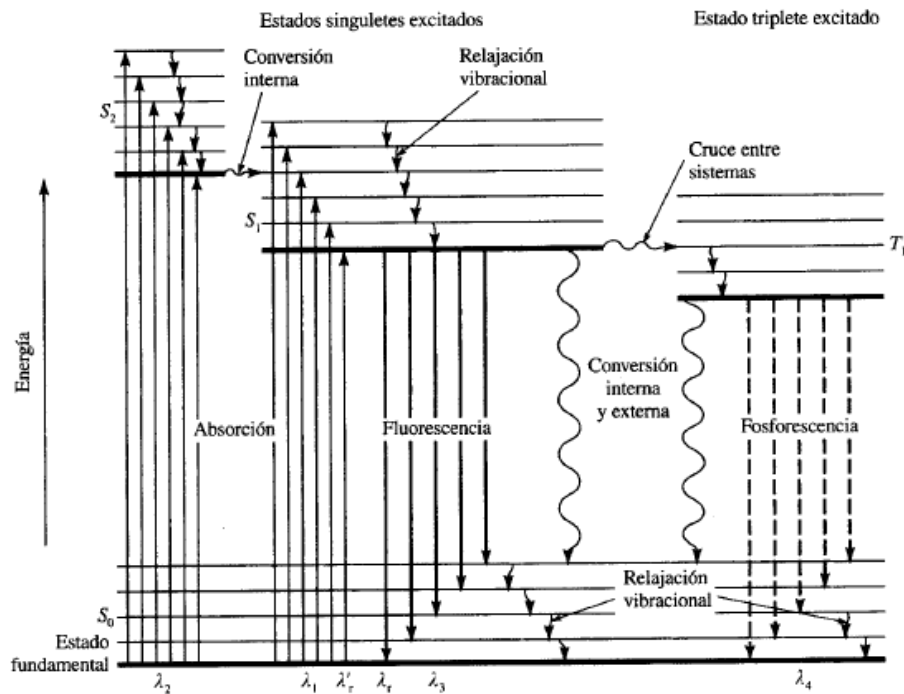


Figura 2. Diagrama de energía parcial para un sistema fotoluminiscente. Fuente: SKOOG, D. A.; HOLLER, F. J. y NIEMAN T. A. Principios de análisis instrumental

La excitación de esta molécula se puede producir por la absorción de dos bandas de radiación, una centrada alrededor de la longitud de onda λ_1 ($S_0 \rightarrow S_1$) y una

segunda alrededor de la longitud de onda más corta λ_2 ($S_0 \rightarrow S_2$); la velocidad de absorción de un fotón es un proceso que requiere del orden de 10^{-15} a 10^{-14} s.

1.2.1. PROCESOS DE DESACTIVACIÓN DE LA FLUORESCENCIA

Sucede que la existencia de la fluorescencia en una sustancia es amenazada por la aparición de otro tipo de fenómenos que logran disminuirla, estos fenómenos están en constante competición con la fluorescencia, y por supuesto estos fenómenos son evidentes en los espectros.⁷

1.2.1.1. Relajación Vibracional

La figura 2, muestra que una molécula puede promocionarse a cualquiera de los diversos niveles vibracionales, sin embargo, en disolución, el exceso de energía vibracional se pierde inmediatamente como consecuencia de las colisiones entre las moléculas de las especies excitadas y las del disolvente, de las cuales resulta un proceso de transferencia de energía¹³. Este proceso de relajación es tan eficaz que el tiempo de vida medio de una molécula excitada vibracionalmente es 10^{-12} s o menos, un periodo significativamente más corto que el tiempo de vida medio de un estado excitado electrónicamente. Es decir, cuando hay fluorescencia de una disolución, siempre se incluye una transición desde el nivel vibracional mas bajo de un estado electrónico excitado, y como consecuencia, la banda de fluorescencia para una transición electrónica dada se desplaza hacia menores frecuencias o longitudes de onda mas largas respecto a la banda de absorción (desplazamiento Stokes).

1.2.1.2. Conversión Interna

Este término describe los procesos intermoleculares por los cuales la molécula pasa a un estado electrónico de menos energía sin emisión de radiación. Según Skoog⁸, estos procesos son evidentemente eficaces, pues hay relativamente pocos compuestos que presentan fluorescencia.

Este proceso tiene lugar cuando dos niveles de energía electrónicos están suficientemente cerca como para que se dé un solapamiento de los niveles de energía vibracional; como sucede con los dos estados singulete excitados en la figura 2, la aparente igualdad de energía potencial hace normalmente más probable esta transición que la fluorescencia desde el estado excitado más alto. Este proceso ocurre a velocidades cercanas a 10^{-12} s.

1.2.1.3. Conversión Externa

La desactivación de un estado electrónico excitado puede incluir la interacción y la transferencia de energía entre la molécula excitada y el disolvente u otros solutos; procesos a los que se les denomina conversión externa o amortiguación colisional. La evidencia de la conversión externa incluye el efecto del disolvente en la intensidad de la fluorescencia, también, aquellas condiciones que tienden a reducir el número de colisiones entre partículas como la baja temperatura y la elevada viscosidad, que tienden, generalmente, a aumentar la fluorescencia.

Según Lakowicz en su libro “principles of fluorescence”²⁰, La idea del solvente debe extenderse no solo a lo que consideramos como solvente, o medio de disolución, sino a los alrededores que acompañan al analito.

1.2.1.4. Cruce entre Sistema

Es un proceso en el cual se invierte el espín de un electrón excitado y da como resultado un cambio en la multiplicidad de la molécula. Igual que en la conversión interna, la probabilidad de esta transición aumenta si los niveles vibracionales de los dos estados se solapan. La figura 2 muestra la transición singulete/triplete como ejemplo de este fenómeno. El cruce entre sistemas es más común en moléculas que contienen átomos pesados como yodo y bromo, y la presencia de oxígeno molecular, que también contribuye a la disminución de la fluorescencia.

1.2.2. VARIABLES QUE AFECTAN LA FLUORESCENCIA

1.2.2.1. La Temperatura

La fluorescencia es altamente sensible a los cambios de temperatura, en ambientes donde hay alta concentración de cromóforos al aumentar la temperatura la intensidad de la fluorescencia disminuye al mismo tiempo que las bandas se hacen más estrechas, la intensidad disminuye principalmente a causa de un apagamiento térmico (la dinámica del sistema) y la estrechez es causada a que los cromóforos son excitados a los estados vibracionales superiores del estado electrónico fundamental.⁵⁴

1.2.2.2. El solvente

Es bien conocido que los solventes que acompañan al fluoróforo pueden generar corrimientos, la introducción de un solvente está relacionado con cambios en la polaridad, constante dieléctrica y cambios en la polarizabilidad del medio que rodea la muestra; los corrimientos por solvente pueden ser contados en términos del efecto sobre todo de las fuerzas de interacción (principalmente del tipo Van der Waals) sobre el sistema π -electrón de la molécula. Este es también conocido que como el sistema π -electrón se vuelve menos localizado, la energía de transición se vuelve más pequeño resultando en corrimiento batocrómico (corrimiento hacia el rojo).^{9,49}

1.2.2.3. El pH

La fluorescencia de un compuesto aromático con sustituyentes ácidos o básicos en el anillo depende normalmente del pH. Tanto la longitud de onda como la intensidad de la emisión son, probablemente, diferentes para las formas ionizada y no ionizada del compuesto. Los cambios en la emisión de los compuestos de este tipo surgen del distinto número de especies resonantes asociadas con las formas

ácidas y básicas de las moléculas. Cuanto mayor es el número de formas resonantes, mayor es la estabilidad del primer estado excitado.

1.2.2.4. Efecto del Oxígeno Disuelto

La presencia oxígeno reduce la intensidad de la fluorescencia de una disolución. Este efecto puede ser el resultado de una oxidación de las especies fluorescentes inducida fotoquímicamente. Con más frecuencia, es la amortiguación que tiene lugar como consecuencia de las propiedades paramagnéticas del oxígeno molecular, que favorece el cruce entre sistemas y la conversión de las moléculas excitadas al estado triplete.

1.2.2.5. Rendimiento Cuántico

Es simplemente la relación entre el número de moléculas que emiten respecto al número total de moléculas excitadas. Algunas sustancias como la fluoresceína, bajo ciertas condiciones, tienen rendimiento cuántico muy cercano a la unidad. Debido a los procesos de desactivación el rendimiento cuántico ϕ para un compuesto se puede calcular a partir de las constantes de velocidad k_x de los procesos por los que el estado singulete excitado más bajo se desactiva.

$$\phi = \frac{k_f}{k_f + k_{ci} + k_{ces} \dots} \quad (\text{Ecuación 1})$$

Donde k_f (fluorescencia), k_{ces} (cruce entre sistemas), k_{ci} (conversión interna), k_{ce} (conversión externa), etc. Es decir, el rendimiento cuántico de emisión de fluorescencia y fosforescencia depende del balance entre las velocidades de relajación radiativa y no radiativa.¹⁰

1.2.2.6. Tipos de Transiciones en Fluorescencia

Rara vez hay fluorescencia como consecuencia de la absorción de radiación UV de longitud de onda menor de 250 nm, ya que esta es suficientemente energética como para producir la desactivación de los estados por pre-disociación o

disociación. Se tiene que radiación de 200 nm corresponde a mas o menos 140 kcal/mol y la mayoría de las moléculas tienen algún enlace que se puede romper con ésta energía. Por ello, rara vez se observa fluorescencia debida a transiciones $\sigma^* \rightarrow \sigma$, en cambio, este tipo de emisión está asociada a los procesos menos energéticos $\pi^* \rightarrow \pi$ y $\pi^* \rightarrow n$.

La fluorescencia surge, generalmente, de una transición desde el nivel vibracional mas bajo del primer estado electrónico excitado a uno de los niveles vibracionales del estado electrónico fundamental. Así, la mayoría de los compuestos fluorescentes, la radiación se produce por una transición $n \rightarrow \pi^*$ y $\pi \rightarrow \pi^*$, dependiendo de cuál de ellas sea menos energética.

1.2.2.7. Fluorescencia y Estructura

La fluorescencia más intensa y más útil es la que presentan los compuestos que contienen grupos funcionales aromáticos con transiciones $\pi \rightarrow \pi^*$ de baja energía. Los compuestos que tienen grupos carbonilo en estructuras alifáticas y alicíclicas o estructuras con dobles enlaces altamente conjugados también pueden presentar fluorescencia, pero el número de estos compuestos es pequeño comparado con el número de sistemas aromáticos. La mayoría de los hidrocarburos aromáticos no sustituidos son fluorescentes en disolución, la eficacia cuántica normalmente aumenta con el número de anillos y con su grado de condensación. Heterociclos sencillos como piridina, furano, tiofeno y pirrol no fluorescen, sin embargo, las estructuras condensadas con estas normalmente sí. Heterociclos con nitrógeno, la transición electrónica de más baja energía probablemente implica un sistema $n \rightarrow \pi^*$ que rápidamente se transforma en un estado triplete e impide la fluorescencia. La sustitución en el anillo bencénico provoca desplazamientos en la longitud de onda de los máximos de absorción y los correspondientes cambios en los picos de fluorescencia al igual que afecta a la eficacia de la fluorescencia. La influencia de la introducción de halógenos causa descenso en la fluorescencia; este descenso, al aumentar el número atómico del halógeno, es debido al efecto

del átomo pesado. La adición de un ácido carboxílico o de un grupo carbonilo en un anillo aromático generalmente inhibe la fluorescencia.^{50,51,53}

1.2.2.8. Efecto de la Rigidez Estructural

Empíricamente se encuentra que la fluorescencia se ve particularmente favorecida en moléculas que poseen estructuras rígidas. La falta de rigidez de una molécula probablemente provoca un aumento de la velocidad de conversión interna y por ello, el aumento en la probabilidad de desactivación no radiante.

1.2.3. FLUORESCENCIA INDUCIDA POR LASER (LIF)

Desde los comienzos de los años 1970, se utilizan diversos tipos de láseres como fuentes de excitación para medidas de foto-luminiscencia. La mayoría de los espectrofluorímetros comerciales utilizan lámparas como fuentes de luz por ser menos caras y menos complicadas en su uso. La ventaja de la inclusión del láser estriba en que es superior a las lámparas en casos como: (1) cuando la cantidad de muestras es muy pequeña, del orden de microlitros o menos, (2) cuando se requiere una radiación de excitación altamente monocromática para minimizar los efectos de las interferencias fluorescentes.

La radiación láser es incluida en los equipos modernos de espectroscopia como fuente de excitación debido a que ésta es monocromática, coherente, y de alta intensidad.³²

Monocromaticidad: se refiere a que la luz en el rayo corresponde a una sola longitud de onda en lugar de una banda de longitudes de onda como en las fuentes de excitación tradicionales.

Coherencia: principalmente hace relación a que el rayo no sufre deformaciones en su morfología, es decir, la sección transversal del rayo no aumenta o disminuye

de tamaño conforme se propaga el rayo, lo que finalmente se traduce en una fuente de excitación muy dirigida y ordenada.

Intensidad: usando un laser se tiene una gran cantidad de fotones impactando contra la superficie en estudio en un tiempo muy pequeño y también sobre un área muy pequeña, es decir, el laser es una fuente de excitación muy eficiente.

Realmente no hay diferencia en la disposición de los montajes para LIF y fluorescencia tradicional, sin embargo, LIF ofrece más posibilidades, atribuibles solo al uso de laser.

1.2.4. MONTAJE PARA LA MEDICIÓN DE LA FLUORESCENCIA

Tradicionalmente la fluorescencia se ha medido a un ángulo de 90 grados respecto a la dirección de la radiación de excitación, hecho que no ha cambiado para la medición de la fluorescencia inducida por laser. Sin embargo, según Noorgaard et al., y en acuerdo a lo dicho en el párrafo anterior, hay un tipo de configuración usada para la medición de fluorescencia en sustancias opacas (que oponen resistencia al paso de la luz a través de ellas) tales como los crudos, esta configuración es denominada modo front-face o modo de cara frontal debido a que la fluorescencia se toma sobre la misma cara de incidencia de la radiación de excitación, a un ángulo $< 90^\circ$ respecto a la dirección del rayo incidente.

1.2.4.1. Configuración Front-Face

Tradicionalmente la fluorescencia es tomada a 90° respecto a la dirección del rayo de incidencia con el objeto de reducir el efecto de llegada de la radiación de excitación al detector en caso que el detector este colocado diametralmente opuesto a la dirección del rayo de incidencia, y para evitar los efectos de reflexión de la radiación de excitación cuando la recolección se hace a ángulos $< 90^\circ$ respecto del rayo de incidencia. La detección a 90° requiere que la muestra este bien diluida, sin embargo, para muestras complejas es frecuente la cuestión de no poder encontrar un disolvente lo suficientemente efectivo y que además garantice

una igual o casi igual capacidad para disolver todos los componentes de una mezcla, para superar esto, se opta por la utilización de la fluorescencia en configuración “front-face” o de cara frontal, que consiste en que la emisión es tomada a un ángulo $< 90^\circ$ respecto al rayo de incidencia; este tipo de configuración experimental ha sido ampliamente reportada en campos como la química analítica de alimentos en los que el espectro de fluorescencia es tomado sobre la superficie del producto, en vinos, pescado, quesos, galletas, etc.^{37,38,39,40,41,42,43,44} convirtiéndola en una herramienta poderosa y amigable con el medio ambiente para hacer estudios de fluorescencia.

1.2.5. FLUORESCENCIA EN CRUDOS

En orden a ser fluorescente, una molécula debería ser altamente absorbente, y poseer una estructura y niveles de energía electrónica apropiados. Para materiales orgánicos, las especies fluorescentes tienden a ser compuestos altamente conjugados no-aromáticos, aromáticos, o heterocíclicos²³. El crudo es fluorescente a causa de que contiene una amplia variedad de especies aromáticas, desde las estructuras más simples (benceno/tolueno) hasta especies muy complejas de alto peso molecular tales como los asfaltenos. La naturaleza de la fluorescencia emitida es grandemente influenciada por la composición química (concentraciones de fluoróforos y apagadores) y características físicas (por ejemplo, viscosidad) del crudo¹². Desafortunadamente, el petróleo comprende un amplio rango de características físicas y químicas, haciendo muy difícil su análisis por fluorescencia⁵².

La longitud de onda de excitación usada para espectroscopia de fluorescencia de crudos debe ser seleccionada con cuidado para permitir una excitación eficiente de todos los tipos de crudos. Diferentes longitudes de onda de excitación resultan en excitación de diferentes poblaciones de fluoróforo, y la complejidad química de los crudos garantiza que en muchos casos haya una alta probabilidad de que la

fluorescencia emitida por una especie sea absorbida por otro fluoróforo resultando en transferencia de energía común en la excitación con UV y azul. Por ello, los estudios que se hacen en petróleo, requieren de geometrías de front-face, o mejor aún, estados altamente diluidos para minimizar estos efectos.

Los procesos de transferencia de energía resultan en corrimiento hacia el rojo y en un ensanchamiento del espectro de emisión, mientras que los procesos de apagamiento reducen la intensidad de la emisión ²². El efecto disminuye con el incremento de la longitud de onda de excitación ¹⁴. Por ejemplo, bajo excitación en el IR cercano, el perfil de espectros de emisión de muestras pesadas, livianas, y diluidas son prácticamente idénticas.

En resumen, los cambios observados en la emisión de fluorescencia con el incremento en la longitud de onda de excitación son: bandas de emisión más estrechas, y reducciones en los corrimientos Stokes, rendimiento cuántico, y tiempos de vida de la fluorescencia.

A pesar de todos estos factores influyendo sobre el fenómeno de la fluorescencia, esta técnica ha servido para empezar a entender el petróleo ^{47,48}, pues la combinación de todos estos fenómenos que acompañan la fluorescencia brindan información adicional sobre este; combinaciones que además implican muchas variables describiendo el comportamiento del sistema químico y que por ello es necesario el uso de sofisticadas herramientas de análisis multivariado como PFA (Principal Factor Analysis), PLS (Partial Least Squares), etc., que han sido usados para analizar la fluorescencia de diversos sistemas orgánicos.⁵⁸

1.3. QUIMIOMETRÍA

Según la sociedad de quimiometría y cualimetría de España ^{15,61}, el término quimiometría fue acuñado entre los 1970 y básicamente resume el concepto que engloba la medida en química, también se refiere a que la quimiometría trata

específicamente de todos aquellos procesos que buscan transformar señales analíticas y datos en información, para lo cual usa métodos de origen matemático, estadístico y los procedentes de la lógica formal.¹⁶

La mayoría de las herramientas quimiométricas ven a un espectro de una sustancia registrado a k longitudes de onda como un vector en un espacio con k dimensiones (p.ej. un espectro tomado a dos longitudes de onda, será un vector en el espacio bidimensional) y de igual modo todo tipo de información descrita por p -parámetros que se puedan saber sobre una muestra también se pueden condensar en un vector, de modo que los espectros y las propiedades de la sustancia obedecerán el algebra vectorial y matricial¹⁷. Si se tienen n muestras cada una de ellas será representada por un punto en el espacio k -dimensional relacionado a la información espectral y también será representada con un punto en el espacio p -dimensional relacionado a la información de propiedades. Así la quimiometría trata de buscar relaciones dentro y entre matrices de información compuestas de estos vectores usando diversas metodologías para poder evidenciar información útil en química^{18,19}. En orden a lo anterior, tenemos dos tipos de vectores, uno espectral y otro de propiedades, y estos se pueden ordenar en matrices espectrales (X) y matrices de propiedades (Y) respectivamente.

La quimiometría comprende una vasta gama de herramientas cada una con finalidad específica, por ello, los métodos quimiométricos podrían clasificarse de acuerdo a su aplicación como sigue:

1.3.1. MÉTODOS PARA EL PRETRATAMIENTO ESPECTRAL

Este tipo de técnicas son empleados para retirar efectos no deseados que acompañan las medidas en espectroscopia conformando lo que se conoce como ruido. Algunos de los pre-tratamientos más usados en espectroscopia son:

Promediado de espectros: Especial para inhibir el efecto del ruido ya que éste es descrito como de naturaleza aleatoria; la reducción de su contribución es lograda

promediando varias señales de una misma muestra aumentando la relación señal/ruido.

Suavizado de espectros: Es usado en aquellos casos donde el promediado no es suficiente. Existen varios algoritmos matemáticos ampliamente reportados en publicaciones y que cumplen a cabalidad con este objetivo; los más habituales son basados en filtros de Savitsky-Golay ²⁴, en transformada de Fourier* ²⁵, y en transformada Wavelet ²⁶.

Transformada de Fourier*: usada cuando se desea descomponer una señal en el dominio del tiempo a otra señal en el dominio de las frecuencias, es decir, en sus diferentes componentes de frecuencia. Usando el concepto de resonancia, una onda de frecuencia variable (llamada fesor u onda de sondeo) es multiplicada punto a punto por la onda que se quiere descomponer (se supone en dominio del tiempo si no lo está), los datos resultantes de este análisis son los valores de aquella multiplicación a cada frecuencia de la onda de sondeo. ²⁶

Derivación de espectros: La utilización de la primera derivada elimina los términos constantes a todas las longitudes de onda, mientras que la segunda derivada corrige las desviaciones causadas por los términos que varían linealmente con la longitud de onda. Suele ser suficiente usar solo la primera y segunda derivadas. La ventaja del uso de la derivada es el aumento de la resolución de bandas, pero como desventaja esta el aumento del ruido. Uno de los algoritmos más usados es el propuesto por Savitsky-Golay, aunque, varios programas comerciales hacen uso de diferencias centrales ²⁷.

1.3.2. MÉTODOS PARA EL PRETRATAMIENTO DE DATOS

Gran parte del éxito de las técnicas robustas de la quimiometría es debida al pre-procesamiento de datos; Edmund Malinowski en su libro Factor Analysis in Chemistry ²⁸, resalta cinco tipos de pre-procesamiento principales. Para asimilar la idea de este autor, es necesario mostrar el pre-procesamiento como un mero proceso de operación entre matrices.

$$X^{\#} = X * A + B \quad (\text{Ecuación 2})$$

Donde una matriz $X^\#$ representa la matriz espectroscópica X de dimensiones $(r \times c)$ que es sometida al pre-tratamiento; A es una matriz diagonal que ajusta todas las magnitudes de cada columna de datos y consiste de elementos a_{jj} ; B es una matriz en la cual todos los elementos b_{ij} en una columna son idénticos, su función es correr el origen del espacio de factores.

Covarianza alrededor del origen, (C_0):

$$a_{jj} = 1 \quad b_{ij} = 0 \quad (\text{Ecuación 3})$$

Covarianza alrededor de la media, (C_m):

$$a_{jj} = 1 \quad b_{ij} = -d_{.j} \quad (\text{Ecuación 4})$$

Correlación alrededor del origen, (R_0):

$$a_{jj} = \left(\sum_{i=1}^r d_{ij}^2 \right)^{-1/2} \quad b_{ij} = 0 \quad (\text{Ecuación 5})$$

Correlación alrededor de la media, (R_m):

$$a_{jj} = \left(\sum_{i=1}^r (d_{ij} - \bar{d}_{.j})^2 \right)^{-1/2} \quad b_{ij} = -\bar{d}_{.j} * a_{jj} \quad (\text{Ecuación 6})$$

Autoscaling a la varianza unidad:

$$d_{ij}(\text{autoescalado}) = \frac{d_{ik} - \bar{d}_{.k}}{s_k} \quad s_k = \left(\frac{1}{r-1} \sum_{i=1}^r (d_{ik} - \bar{d}_{.k})^2 \right)^{1/2} \quad (\text{Ecuación 7})$$

Autoscaling a longitud unidad:

$$d_{ij}(\text{autoescalado}) = \frac{d_{ik} - \bar{d}_{.k}}{\|d_{.j}\|} \quad (\text{Ecuación 8})$$

1.3.3. MÉTODOS PARA LA REDUCCIÓN DIMENSIONAL

Debido a la gran cantidad de datos ligados a la espectroscopia, surge la necesidad de reducir el número de variables con las que se va a trabajar y así es como

aparece una técnica ampliamente usada en trabajos de diversas disciplinas llamada PFA (Principal Factor Analysis).^{29, 30}

1.3.3.1. Análisis de Factores Principales (PFA)

O mejor conocido como PCA (análisis de componentes principales) tiene dos funciones principales, la reducción dimensional y poner en evidencia algún tipo de comportamiento (Pattern Recognition). Esta técnica consiste en proyectar los espectros en un nuevo sistema coordinado de menos dimensiones descrito por ejes ortogonales que son llamados factores principales.

La primera de sus funciones se consigue hallando las direcciones que explican la máxima variabilidad de las muestras y usarlas como nuevos ejes de coordenados a los cuales se les conoce como Factores Principales (PFs). Se ha referido al PFA como un proceso de rotación de los ejes coordinados originales, sin embargo, ya que muchas de las variables originales son redundantes o carecen de información valiosa, el sistema puede representarse con un menor número de ejes que el original, de esta forma se reduce la Dimensionalidad. El primer factor principal (PF1) es una combinación lineal de las k variables que explica la máxima variabilidad de las muestras. El segundo PF (PF2) se escoge de forma que sea ortogonal a primero y que explique la máxima variabilidad de las muestras una vez restada la explicada por PF1. Estos nuevos ejes son definidos por los loadings que son los cosenos de los ángulos que forman los nuevos ejes con los originales, y los scores que son las coordenadas de las muestras en los nuevos ejes.

Numéricamente, la matriz X (espectroscópica) es descompuesta en un producto de dos matrices: una matriz de scores T y una matriz de loadings P, con un residual condensado en la matriz E.

$$X = T * P^t + E \quad (\text{Ecuación 9})$$

Cada factor principal contiene información de diferente relevancia. PF1 contiene la mayor variabilidad de los datos, PF2 tiene la segunda mayor variabilidad en los datos, y así sucesivamente hasta contener toda la información importante. La

representación completa de la matriz X implica tantos componentes (tantos vectores de loadings y scores) como dimensiones k tengan los espectros. Sin embargo, uno de los objetivos de PFA es la reducción de la Dimensionalidad, la matriz original será representada por un número menor de vectores $a < k$.

$$X = t_1 * p_1^t + t_2 * p_2^t + \dots + t_a * p_a^t + E \quad (\text{Ecuación 10})$$

$$X = T_a * P_a^t + E = X_a + E \quad (\text{Ecuación 11})$$

Y así el conjunto de datos X, que inicialmente estaba descrito por variables correlacionadas, queda definido por un nuevo conjunto de variables (PFs) no correlacionadas entre sí en un nuevo sistema de ejes ortogonales.

Un ejemplo de cómo funciona es mostrado en la figura 3, donde ocho muestras tienen espectros tomados a dos longitudes de onda. El sistema puede ser reordenado respecto a dos nuevos ejes, sin embargo, uno de ellos recoge la mayoría de la variabilidad en las muestras razón por la cual el sistema puede ser representado únicamente por PF1.

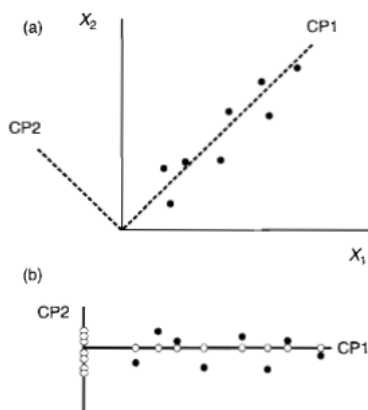


Figura 3. Esquema de funcionamiento del PFA. Fuente: MILLER, J. Estadística para Química Analítica.

1.3.3.2. Mínimos Cuadrados Parciales (PLS)

Una vez centradas o auto-escaladas las matrices X e Y, cada una de las matrices se descompone en una suma de A factores calculados simultáneamente:

$$X = T * P^t + E = \sum_{a=1}^A t_a * p_a^t + E \quad (\text{Ecuación 12})$$

$$X = X_A + E; \quad (\text{Ecuación 13})$$

$$Y = U * Q^t + F = \sum_{a=1}^A u_a * q_a^t + F \quad (\text{Ecuación 14})$$

$$Y = Y_A + F; \quad (\text{Ecuación 15})$$

Algo muy importante es que los loadings de X calculados por PFA no coinciden con los calculados por PLS ya que en el caso de este último los loadings están corregidos para obtener la máxima capacidad predictiva para Y. ^{57,58}

1.3.4. MÉTODOS PARA EL RECONOCIMIENTO DE PATRONES

La quimiometría tiene su fortaleza en la solución o análisis de sistemas con una alta complejidad, así que clasificar las muestras puede en principio ayudar a disminuir la complejidad del sistema que se está estudiando. En ciencias así como en ingeniería es deseable tener un criterio matemático que permita asignar una muestra nueva a uno de los grupos en que se ha dividido la población de muestras. El aspecto clave aquí es identificar las variables discriminantes y su respectiva característica discriminada, esta última, en la mayoría de los casos corresponde a una característica cualitativa (por ejemplo, la característica discriminada puede ser que tan buen atleta es una persona, una variable discriminante debería ser la velocidad obtenida en pruebas de pista). Una de las metodologías más usadas por los quimiometristas es el LDA (Linear Discriminant Analysis), que depende en muchos casos de la separación de muestras luego de la reducción dimensional, en otros casos, en donde no es necesaria la reducción para apreciar el pattern recognition, como cuando se posee un amplio conocimiento de la muestra así como del fenómeno medido y de la morfología de

los espectros, la identificación de las variables es más simple y se hace directamente de los espectros originales, a estas se les denomina variables principales.

1.3.4.1. Análisis Discriminante Lineal (LDA)

Es una técnica de reconocimiento de pautas asistido, es decir, que la relación entre un grupo de variables que permiten separar una población de medidas en algunos sub-grupos y la característica que los hace diferentes entre sí, es asignada o buscada por el analista; en otras palabras una vez encontradas las variables que permiten distinguir cierta particularidad entre las muestras y se ha logrado distinguir la propiedad que hace diferentes los grupos, LDA sirve como medio para obtener una fórmula matemática para describir tal separación. Dada la gran utilidad de LDA, a través de los años se han hecho mejoras a los procedimientos matemáticos ³³, incluso hasta tiene además de la versión lineal, versión cuadrática, etc. ^{34,55}, pero el uso de una o de otra de sus formas depende de la disposición que tengan los datos; en la siguiente figura x_1 y x_2 son la variables discriminantes porque permiten una separación de los datos en puntos blancos y puntos negros donde la característica discriminatoria es el color, como resultado de aplicar LDA se obtiene una función $Y(x_1, x_2)$ lineal ya que los grupos tienen una dispersión alargada que se puede ajustar a un comportamiento lineal.

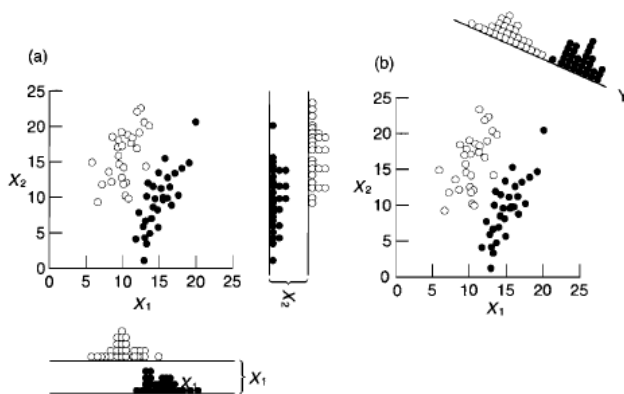


Figura 4. Esquema de funcionamiento de LDA. Fuente: MILLER, J. Estadística para Química Analítica.

La forma como procede LDA es calculando las distancias entre muestras o entre vectores, para ello usa principalmente dos maneras de ver estas distancias: distancia Euclidiana y distancia de Mahalanobis.

Esta técnica en asocio con PFA ha sido usada con buenos resultados en análisis de gasolinas.^{45, 59}

1.3.5. MÉTODOS PARA LA CONSTRUCCIÓN DE MODELOS PREDICTIVOS

Una vez clasificadas las muestras, uno de los objetivos generales de la quimiometría -la Construcción de modelos predictivos -, es posible gracias a varios métodos; el objetivo de estos es encontrar la forma como se relacionan dos matrices de datos, en este caso, una espectroscópica y una de propiedades.

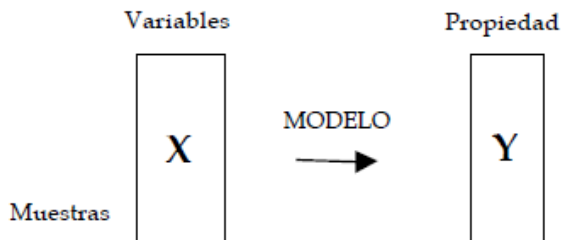


Figura 5. Esquema fundamental de la quimiometría. Fuente:
<http://www.quimica.urv.es/quimio>

El anterior esquema responde a la siguiente ecuación matricial

$$X * B = Y \quad (\text{Ecuación 16})$$

Donde la única incógnita es B o matriz regresora, pero es determinada como:

$$B = (X^t * X)^{-1} * X^t * Y \quad (\text{Ecuación 17})$$

Principalmente tres técnicas son usadas para esta finalidad y lo único que las diferencia entre sí son las variables que componen la matriz espectroscópica o X en la figura anterior; estas técnicas son: MLR (Multiple Linear Regression), PCR (Principals Components Regression), y PLSR (Partial Least Squares Regression).

1.3.5.1. Regresión Lineal Múltiple (MLR)

Es útil cuando no es necesario emplear técnicas de proyección ortogonal como PFA, en lugar de ello, se escogen las variables que el analista considera son de relevancia para sus estudios, llamadas variables principales, Noorgard et al ³⁵, entonces, la matriz espectroscópica ya no es (m x k), sino (m x n) con n < k. La fórmula anterior cambia de apariencia a la siguiente:

$$Y = X_n * B \quad (\text{Ecuación 18})$$

Donde la matriz de regresores se calcula como:

$$\ddot{B} = (X_n^t * X_n)^{-1} * X_n^t * Y \quad (\text{Ecuación 19})$$

Un grupo de medidas nuevas serán sometidas a los pre-tratamientos del grupo de calibración, se escogen las mismas VPs, y se usa la (ecuación 16):

$$Y^* = X_n^* * \ddot{B} \quad (\text{Ecuación 20})$$

El problema de esta técnica es que si no se escogen correctamente las variables, la calibración falla debido a que se pueden generar singularidades, es decir, que en la (ecuación 19), el determinante de $(X_n^t * X_n)$ es igual o muy cercano a cero.

1.3.5.2. Regresión por Componentes Principales (PCR)

Se usa cuando no se tiene mucho conocimiento sobre el sistema que se está analizando y el único recurso que queda para escoger variables que contengan información química es la reducción dimensional. El uso de PCR se hace útil cuando los ejes coordenados o factores calculados por PFA tienen una buena correlación con la matriz de propiedades, de esta forma la matriz espectroscópica estará descrita por las puntuaciones (scores) de las muestras en los nuevos ejes en lugar de las variables originales.

El primer paso en PCR consiste en descomponer la matriz X de acuerdo a la (ecuación 11) y una vez se ha considerado A como el número óptimo de factores. La matriz espectroscópica se puede representar por su matriz de puntuaciones T

$$T = X_A * P \quad (\text{Ecuación 21})$$

Entonces la matriz de datos se puede calcular según:

$$Y = T * B + E \quad (\text{Ecuación 22})$$

Entonces,

$$\hat{B} = (T^t * T)^{-1} * T^t * Y \quad (\text{Ecuación 23})$$

Donde el símbolo $\hat{}$ indica valores calculados. Y ahora para realizar los cálculos para predecir un conjunto de nuevas muestras X^* (muestras de predicción), se debe aplicar el mismo proceso de pre-tratamiento que a los datos de calibración. A partir de las cargas (loadings) de calibración, se calculan las puntuaciones de las nuevas muestras.

$$T^* = X_A^* * P \quad (\text{Ecuación 24})$$

Y usando esta matriz junto con la matriz de regresores, se calcula la matriz propiedades para las muestras desconocidas

$$Y^* = T^* * \hat{B} \quad (\text{Ecuación 25})$$

El problema con esta técnica es que los factores principales que mejor representan la matriz de los datos espectroscópicos no logran ser los más convenientes para la predicción de las propiedades que se están estudiando,

entonces se les aprovecha para observar algún tipo de comportamiento como agrupamientos, etc., aprovechando la posibilidad gráfica que proporciona la reducción dimensional.

1.3.5.3. Regresión por Mínimos Cuadrados Parciales (PLSR)

Esta técnica también es útil donde el analista tiene poca información de cómo responde el sistema a la estimulación instrumental, pero presenta una notable ventaja sobre PCR, y es que a diferencia de este, PLS tiene en cuenta que en la mayoría de los casos las dos matrices tienen diferente dimensionalidad por lo cual calcula otro tipo de variables conocidas como factores latentes que son ejes comunes a ambas matrices, la espectroscópica y la de propiedades; la idea es similar a PFA solo que en PLSR se proyectan las dos matrices en un sub-espacio constituido por unos nuevos ejes coordenados o factores.

El primer paso en PLSR consiste en descomponer simultáneamente la matriz X e Y de acuerdo a las ecuaciones 12 y 14, estableciendo una relación interna entre las puntuaciones de los datos X e Y de forma tal que para cada factor a se cumpla:

$$\hat{u}_a = b_a * t_a \quad (\text{Ecuación 26})$$

Donde b_a es el coeficiente de regresión para cada uno de los factores latentes.

Una vez se ha considerado A como el número óptimo de factores. La matriz espectroscópica y de propiedades se relacionan a través de sus matrices de puntuaciones:

$$\hat{U} = B * T \quad (\text{Ecuación 27})$$

Y ahora para realizar los cálculos para predecir un conjunto de nuevas muestras X^* (muestras de predicción), se debe aplicar el mismo proceso de pre-tratamiento que a los datos de calibración. A partir de las cargas de calibración, se calculan las puntuaciones de las nuevas muestras.

$$T^* = X_A^* * P \quad (\text{Ecuación 28})$$

Y a través de la (ecuación 27) se calculan las puntuaciones de la matriz propiedad de las muestras desconocidas:

$$\bar{U}^* = B * T^* \quad (\text{Ecuación 29})$$

Y usando las cargas de la matriz propiedad se calcula la matriz Y para las nuevas muestras:

$$Y_A^* = \bar{U}^* * Q^t \quad (\text{Ecuación 30})$$

PLS tiene dos formas: PLS1 es el algoritmo encargado de calcular una única propiedad, y PLS2 es el algoritmo encargado de calcular varias propiedades.

1.3.6. MÉTODOS PARA EVALUAR LA CAPACIDAD PREDICTIVA

La quimiometría también comprende las metodologías que se usan para determinar si un modelo es válido o no. La literatura muestra principalmente tres figuras de merito que se usan para tal fin; (1) la RMSEC (Root Mean Square Error Calibration) o raíz cuadrada del error cuadrado medio de calibración, (2) la RMSEP (Root Mean Square Error Prediction) o raíz cuadrada del error cuadrado medio de predicción, también conocida como la RMSEV (Root Mean Square Error Validation) o raíz cuadrada del error cuadrado medio de validación, y (3) la RMSECV (Root Mean Square Error Cross Validation) o raíz cuadrada del error cuadrado medio de calibración cruzada.³²

- (1) Describe el grado de acuerdo entre los valores de concentración estimadas por el modelo de calibración para las muestras de calibración y los valores aceptados como verdaderos para las muestras de calibración usadas para obtener los parámetros del modelo:

$$RMSEC = \left(\frac{\sum (\hat{y}_{ij} - y_{ij})^2}{m - 1 - A} \right)^{1/2} \quad (\text{Ecuación 31})$$

- (2) Las muestras preparadas y medidas independientemente de las muestras de calibración son usadas. El numero de muestras de validación, p, debería

ser grande, tanto que el error de predicción estimado refleje exactamente todas las fuentes de variabilidad en el método de calibración:

$$RMSEP = \left(\frac{\sum (\hat{y}_{ij} - y_{ij})^2}{m_p} \right)^{1/2} \quad (\text{Ecuación 32})$$

- (3) La calibración cruzada se usa para estimar la habilidad predictiva de un modelo de calibración. Uno de los métodos es el de dejando uno por fuera (LOOCV), que es desarrollado estimando n modelos de calibración, donde cada uno de las n muestras de calibración es dejada afuera, una por una hasta que todas han tenido su turno; y los modelos resultantes son usados para estimar la muestra que se ha dejado por fuera.

$$RMSECV = \left(\frac{\sum_{i=1}^m (\hat{y}_{ij} - y_{ij})^2}{m} \right)^{1/2} \quad (\text{Ecuación 33})$$

2. PARTE EXPERIMENTAL

Este trabajo investigativo tiene lugar bajo convenio de cooperación tecnológica Universidad Industrial de Santander – Instituto Colombiano del Petróleo UIS-ICP.

2.1. MONTAJE EXPERIMENTAL

El montaje experimental se realizó como se muestra en la figura 6. Como fuente de excitación se utilizó un láser pulsado de Nitrógeno trabajando a una frecuencia de 4.5 Hz. La radiación proveniente del láser se enfocó hacia la muestra (3) localizada a una distancia focal de 15 cm, a través de un lente cilíndrico (2). La radiación de excitación de la muestra se recolectó por medio de una fibra óptica (4). Esta fibra se encargó de conducir la radiación al espectrómetro (5), y esta fue llevada a un PC (6) para su análisis

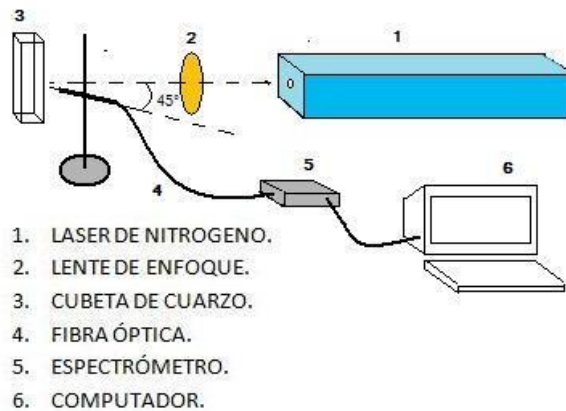


Figura 6. Esquema experimental de LIF front-face.

- (1) Láser pulsado de Nitrógeno PL 2300
- (2) Lente cilíndrico
- (3) Porta muestra

- (4) Fibra Óptica
- (5) Espectrómetro
- (6) Ordenador o PC

En este trabajo se partió de esta configuración básica, proponiendo y explorando diferentes disposiciones hasta encontrar que la señal de la fluorescencia es mayormente visible situando la fibra óptica a 45° respecto el rayo laser de excitación (ver figura 6).

2.1.1. COMPONENTES DEL MONTAJE EXPERIMENTAL

El sistema LIF mostrado en la figura, consta de los siguientes componentes:

2.1.1.1. Láser de Nitrógeno

Se utilizó un laser pulsado de nitrógeno (Modelo PL2300/PL202 PTI), (*PTI Laser INC*) (Figura 7) como fuente de excitación, con una longitud de onda de 337.1 nm (337 nm), 1.4 mJ por pulso, potencia 2.3 MW, ancho temporal de 600ps, ancho de banda 0.1 nm y frecuencia de 20 Hz.



Figura 7. Consola del Láser de nitrógeno PL 2300

2.1.1.2. Lente cilíndrico

Se utilizó un lente cilíndrico (Figura 8) para enfocar la radiación hacia el sitio específico, el cual presenta una de las caras cilíndricas, en donde al penetrar un haz de luz a su través en el punto de enfoque aparecerá una línea.

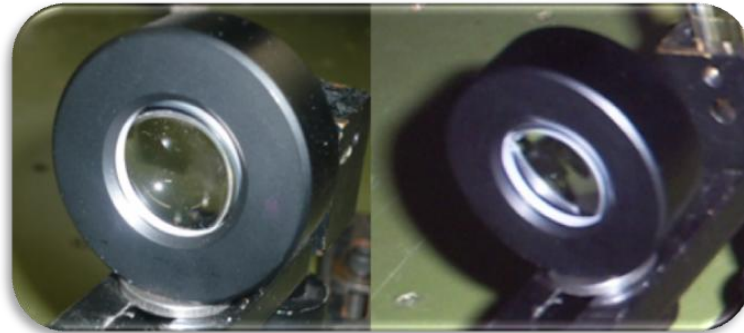


Figura 8. Lente cilíndrico

2.1.1.3. Fibra óptica

Fue usada una fibra óptica (figura 9) con núcleo de sílice pura, diámetro de 50 μm , y rango de UV-VIS (200-1100 nm), como el medio de transmisión de la luz emitida hacia el espectrómetro.

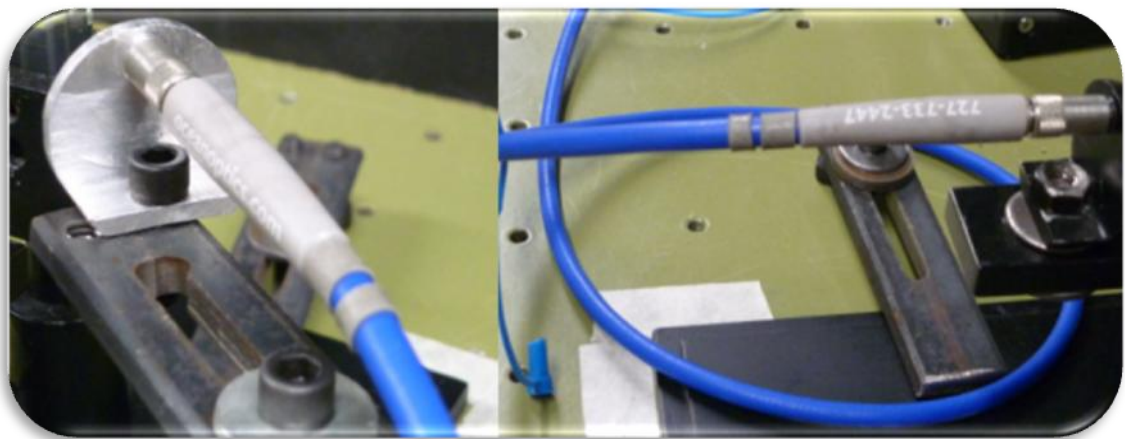


Figura 9. Fibra óptica

2.1.1.4. Espectrómetro (USB4000 Ocean Optics)

Se utilizó un espectrómetro USB 4000 de Ocean Optics (Figura 10), con una matriz CCD lineal Toshiba de 3648 receptores, con un ancho espectral de 350-1000 nm y un tiempo de integración de 10 μs ; este dispositivo mide la intensidad

de la luz dispersada sobre la matriz de receptores en su interior cada uno de los cuales muestra la intensidad a una longitud de onda determinada.



Figura 10. Espectrómetro (USB4000 Ocean Optics)

2.1.1.5. Porta Muestra

Como porta muestra se utilizó una celda de cuarzo (Figura 11) con un paso de luz de 1cm.

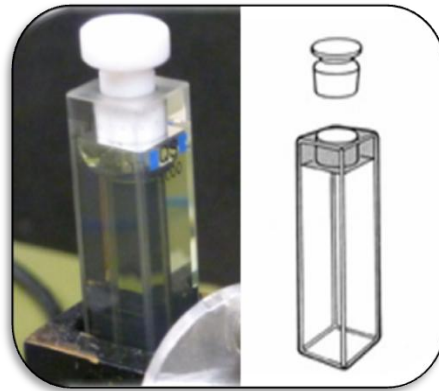


Figura 11. Celda de 1 cm

2.2. ADQUISICIÓN DE DATOS

Los estímulos que luz que llegan a los receptores del espectrómetro son convertidos en tensiones eléctricas y después digitalizadas por medio de un software de control llamado SpectraSuite Ocean Optics.

Luego de registrar un espectro de prueba para el crudo C1 y previendo que se encontraba en acuerdo con los mostrados en la literatura¹¹ (figura 13), fueron registrados los espectros de cada uno de los 20 crudos.

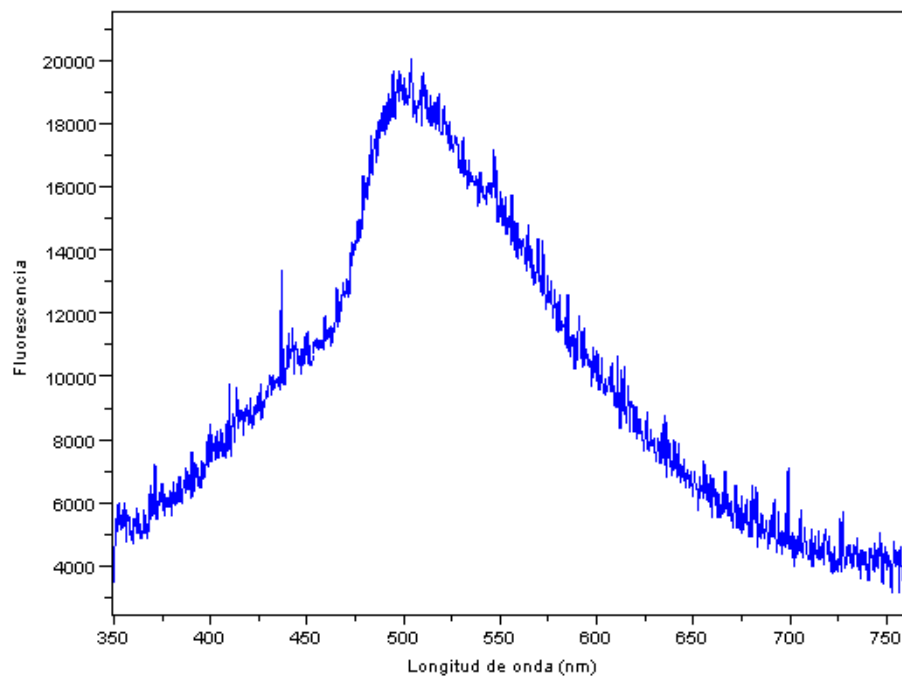


Figura 12. Espectro LIF del crudo C1

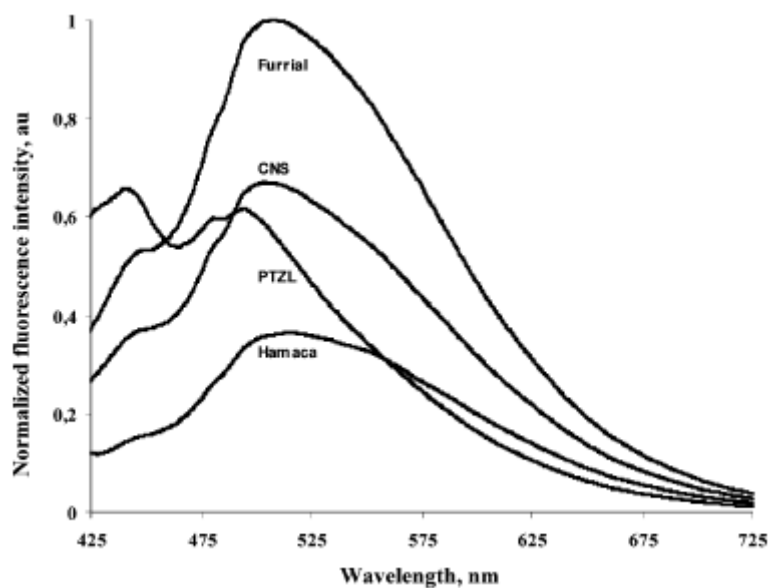


Figura 13. Espectros encontrados en la literatura

Una vez verificada la similitud, se registran los espectros de los correspondientes 20 crudos.

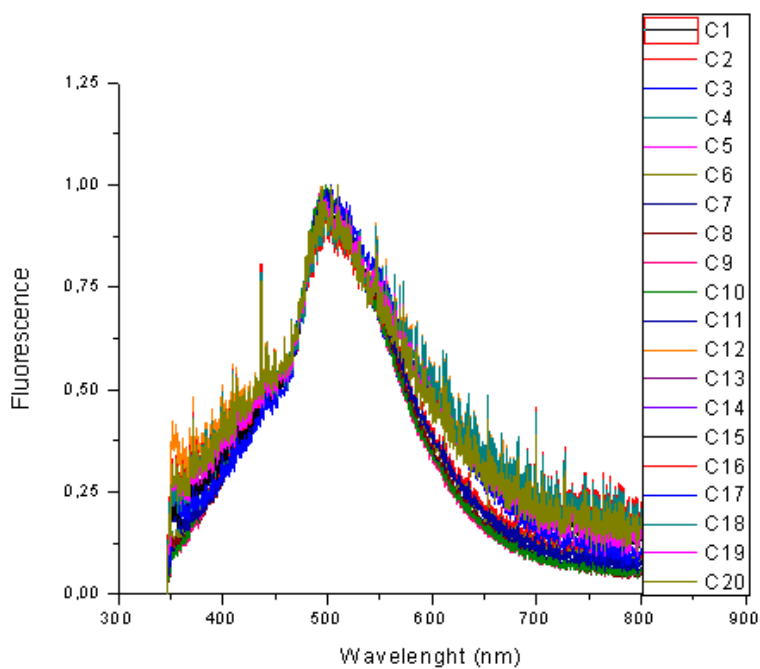


Figura 14. Espectros LIF de los 20 crudos .

2.3. PROCESAMIENTO DE DATOS

Una vez obtenidos todos los espectros, el análisis quimiométrico se realizó usando Scilab 5.2.2,⁶⁰ una plataforma de programación de licencia libre y muy similar a MATLAB, aunque tiene herramientas estadísticas y de análisis multivariado, es necesaria una programación de las funciones de tal forma que cumpla un número de funciones definidas y de forma ordenada sobre los datos (ver anexos).

Previo al análisis, es necesario un pre-tratamiento consistente en hacer un alisado usando un filtro Savitsky-Golay (usando un polinomio de segundo orden, en una ventana de 11 datos), luego del cual se retira la región espectral de baja relación señal/ruido, dejando únicamente el intervalo de 350 nm – 800 nm.

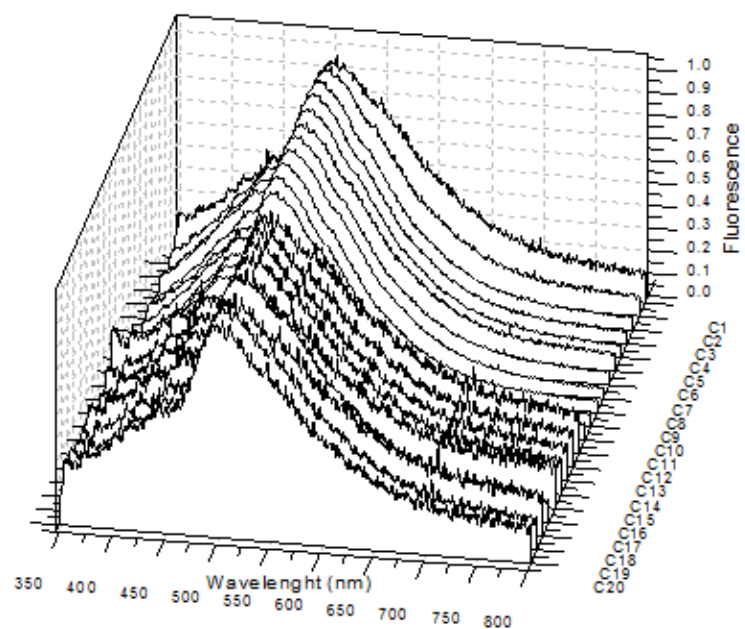


Figura 15. Espectros LIF de los 20 crudos después de alisamiento.

También se le aplicó la corrección de la línea base (Figura 16) con el objetivo de eliminar la señal de ruido que puede aportar la línea base a cada uno de los espectros.

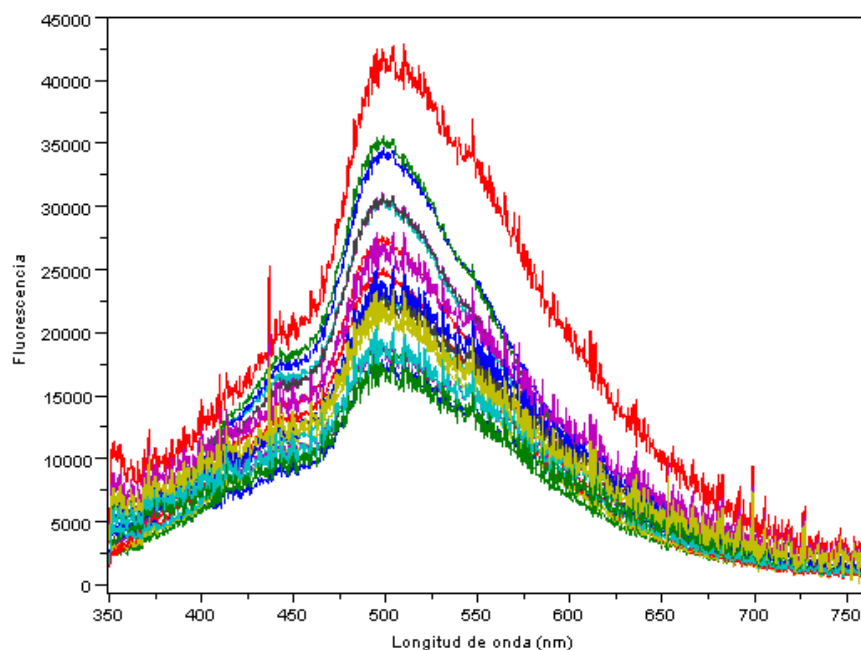


Figura 16. Corrección de línea base.

Otro de los pretratamientos aplicados fué la normalización a la longitud unidad (figura 17); el objetivo de la normalización es estandarizar la medida, es decir, la figura anterior muestra que a pesar que las medidas se hicieron bajo las mismas condiciones experimentales, su apariencia es diferente, principalmente en lo que respecta a la intensidad y a la anchura de los espectros, y por ello se escalan las mediciones a un valor referencia propio de cada muestra, en este caso, la magnitud del espectro (visto como un vector); la práctica de normalizar las mediciones es algo que se puede considerar característico en algunas técnicas como la espectrometría de masas.

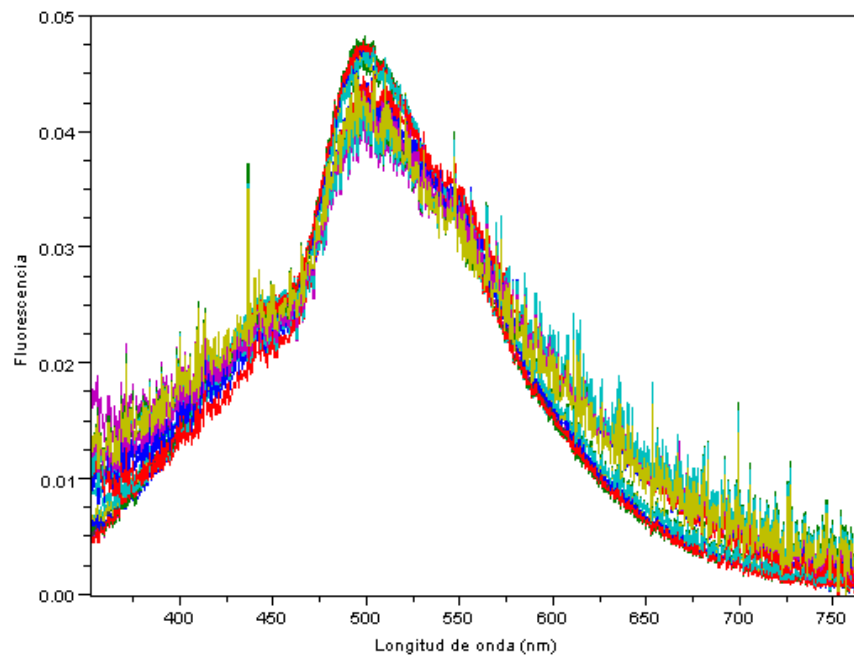


Figura 17. Espectros normalizados a longitud unidad

Después de realizar cada uno de los pretratamientos se realizó el análisis de PFA, para poder evidenciar algún tipo de comportamiento en los datos como conglomerado, luego de lo cual se usa LDA para describir matemáticamente el comportamiento de las muestras como un grupo, y finalmente, se usa MLR para construir los modelos de predicción de la fracción S.A.R.A.

3. RESULTADOS Y ANÁLISIS

A continuación se presentan de una forma resumida los principales resultados obtenidos en los análisis de factores principales (PFA), LDA, MLR, para los crudos colombianos analizados.

3.1. ANÁLISIS DE FACTORES PRINCIPALES (PFA)

Como se advirtió en la sección 1.3.3.1., PFA tiene la capacidad no solo de reducir las dimensiones del espacio de las variables sino una posibilidad de evaluación gráfica de las muestras en dicho espacio reducido permitiendo con esto evidenciar algún tipo de reconocimiento de patrones (pattern recognition).

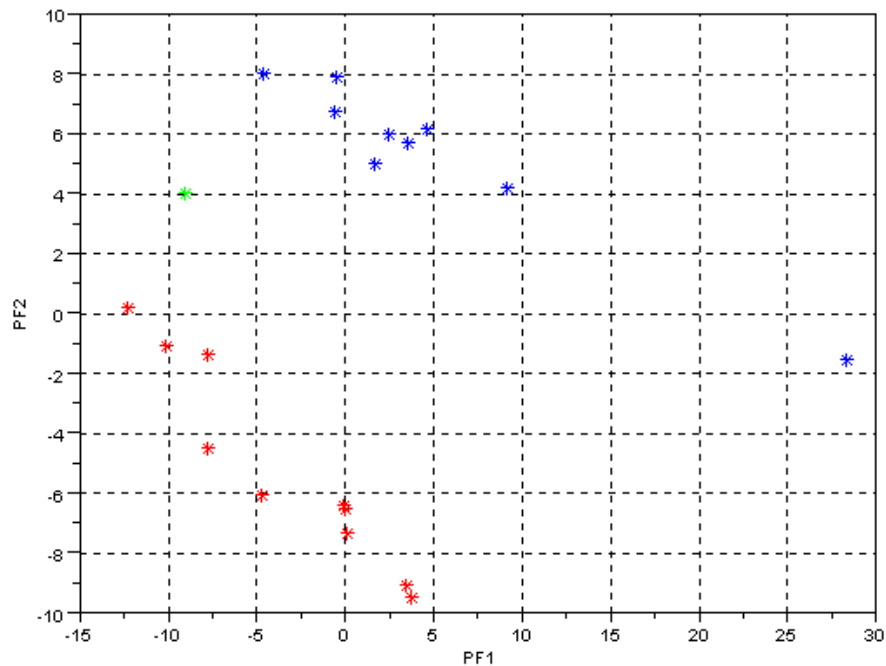


Figura 18. Resultados PFA. PF1 vs. PF2

La figura 18 muestra que como resultado de hacer PFA sobre la matriz espectroscópica con variables normalizadas alrededor de la media, y graficar los scores del PF1 (Factor Principal 1) vs los scores de PF2 (Factor Principal 2), hay

una evidente separación de las muestras en un grupo de color rojo, otro de color azul, y una muestra representada por el asterisco verde. Esta separación coincide exactamente con que las muestras provienen de tres campos petroleros (por petición del Instituto Colombiano del Petróleo el nombre y origen de las muestras no es revelado).

3.2. ANÁLISIS DISCRIMINANTE LINEAL (LDA)

Una vez lo evidenciado en la figura anterior, una forma matemática de describir esta separación, se obtiene de la aplicación de LDA a las muestras en términos únicamente de PF1 y PF2. Al no haber sino una muestra perteneciente al grupo representado por el color verde, la función discriminante se construyó teniendo en cuenta solo las muestras pertenecientes a los otros dos grupos. Así es como la función discriminante viene dada por la (ecuación 34):

$$W = (-3.4405086) * PF1 + (-8.8089957) * PF2 + (-2.9857641) \quad (\text{Ecuación 34})$$

Para la evaluación de una muestra desconocida, se reemplazan en la ecuación anterior los valores de sus scores en PF1 y PF2; la función discriminante funciona de la siguiente manera: si $W > 0$, la nueva muestra pertenece al grupo rojo (grupo 1) que corresponde a las muestras C2 – C11, mientras que si $W < 0$, entonces la muestra pertenece al grupo azul (grupo 2) que corresponde a las muestras C12 – C20.

3.3. CONSTRUCCIÓN DEL MODELO PREDICTIVO (MLR)

Uno de los problemas más ampliamente discutidos sobre los métodos de calibración multivariable es cuantas y cuáles de las nuevas variables (variables principales, factores principales, y variables o factores latentes principales), se deben tener en cuenta para la construcción de un buen modelo que prediga la propiedad deseada, en este caso, la fracción S.A.R.A. de los 20 crudos (ver tabla 1); los criterios son muchos, sin embargo, todos apuntan a obtener el error de predicción más pequeño.⁴⁶

Noorgard *et al*³⁵. Muestra que una de las estrategias para encontrar las variables principales es hacer regresiones lineales entre cada variable espectroscópica (datos de una sola posición espectral para todos las muestras) contra cada variable de propiedades (datos de una sola variable para todos las muestras), y calcular los errores involucrados para cada posición espectral, haciendo esto para cada una de las 4 variables: saturados, aromáticos, resinas, y asfaltenos, se obtienen unos gráficos (figuras 19 a 22) en donde la ordenada representa un parámetro llamado RSSEy o raíz cuadrada de la suma de los errores cuadrados en la variable de propiedades (y). El uso de metodologías para determinar regiones de interés no es propio de MLR únicamente, de hecho, en técnicas de proyección ortogonal como PLS también es usado³⁶.

La tabla 1 contiene la información de la fracción S.A.R.A., para los 20 crudos, medida mediante los protocolos ASTM (figura 1).

Tabla 1. Fracción S.A.R.A. para los 20 crudos (valores referencia)

Espectros	SATURADOS	AROMATICOS	RESINAS	ASFALTENOS
C1	33.57	30.56	33.57	2.3
C2	42.5	30	26.88	0.62
C3	46.82	28.09	24.75	0.34
C4	45.22	25.03	29.07	0.68
C5	38.4	32.49	28.8	0.32
C6	41.63	24.64	33.13	0.6
C7	43.89	31.46	24.14	0.51
C8	38.5	32.53	28.55	0.42
C9	46.66	24.52	28.47	0.36
C10	46.61	26.99	26.17	0.23
C11	39.41	29.56	30.32	0.72
C12	36.73	34.79	26.52	1.96
C13	39.18	30.53	24.42	5.87
C14	30.71	31.94	33.78	3.57
C15	36.81	35.18	24.9	3.11
C16	38.12	34.83	24.95	2.1
C17	37.15	30.96	30.4	1.49
C18	40.02	30.26	26.85	2.87
C19	35.77	33.73	27.09	3.41
C20	42.55	28.21	25.44	3.8

Una de las observaciones que se deben hacer previo cualquier análisis es ver la forma que tiene o tienen las variables que se desean analizar; de la tabla 1 se ve que a pesar de haber dos grupos establecidos por PFA y definidos por LDA, dentro de estos hay grandes variaciones de los valores de S.A.R.A., es decir, aunque dos muestras pertenezcan a un mismo campo petrolero, dichas muestras pueden diferir notablemente como por ejemplo, los crudos C19 y C20; debido a esta consideración, y que la población de muestras es pequeña, el modelo se construye sobre la totalidad de estas ya que los datos de la tabla 1 muestran que en los dos grupos es común que hayan tanto valores altos como bajos de todas las fracciones, es decir, omitir uno, limita o priva al modelo de información vital, como si cada una de todas las muestras aportaran información necesaria para describir el comportamiento del sistema.

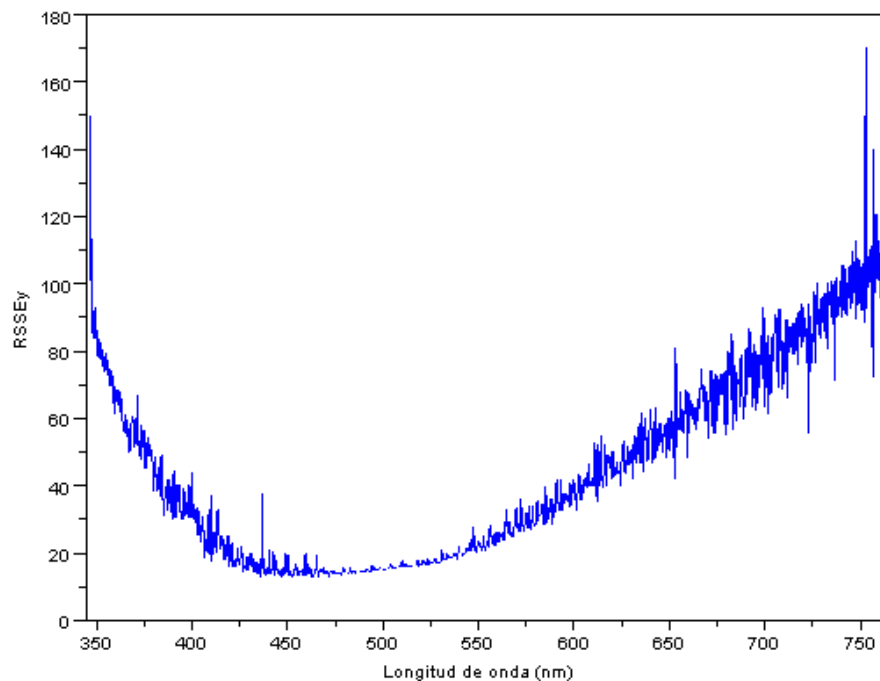


Figura 19. Regresiones locales para saturados

La figura 19 muestra que entre 425 nm y 550 nm, se encuentra una zona donde el error de la regresión es mínimo.

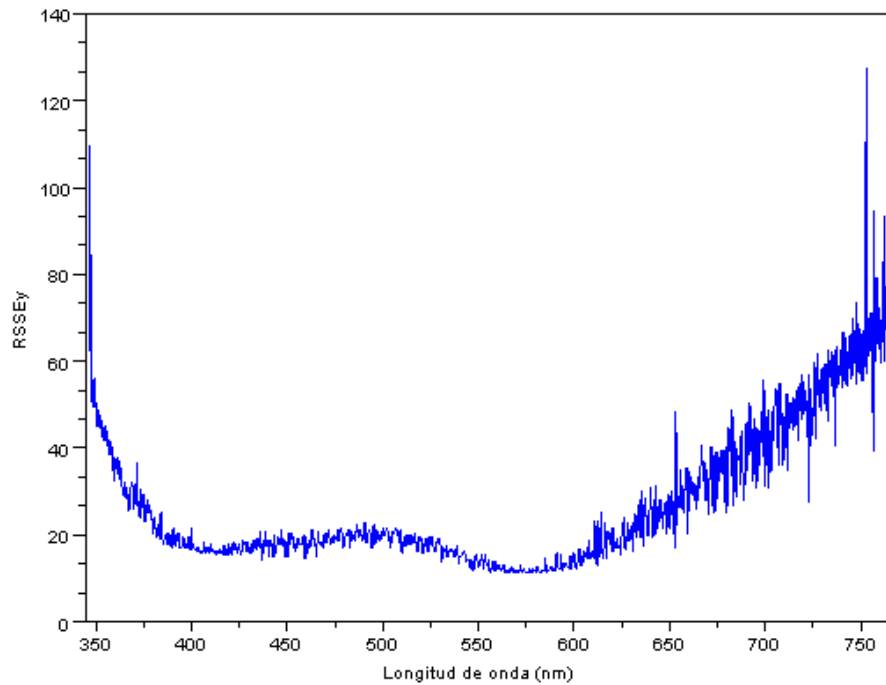


Figura 20. Regresiones locales para aromáticos

La figura 20 muestra que entre 375 nm y 625 nm, pero con un énfasis especial en 400 nm y en 575 nm, se encuentra una zona donde los errores de regresión son mas bajos respecto al resto de la curva

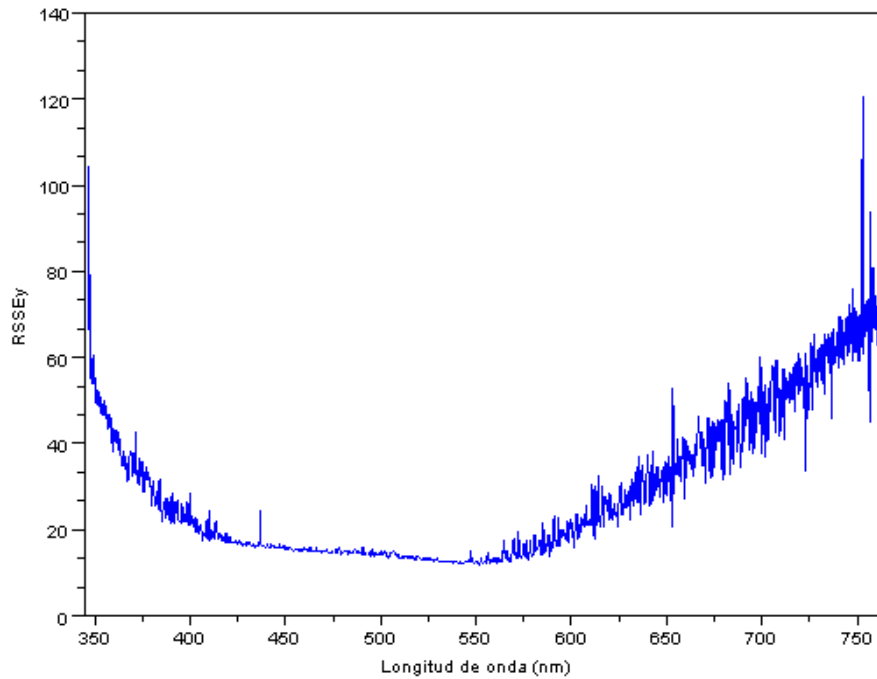


Figura 21. Regresiones locales para resinas

La figura 21 muestra que entre 400 nm y 575 nm, pero con un énfasis especial en 425 nm y en 550 nm, se encuentra una zona donde los errores de regresión son menores que los del resto de la curva.

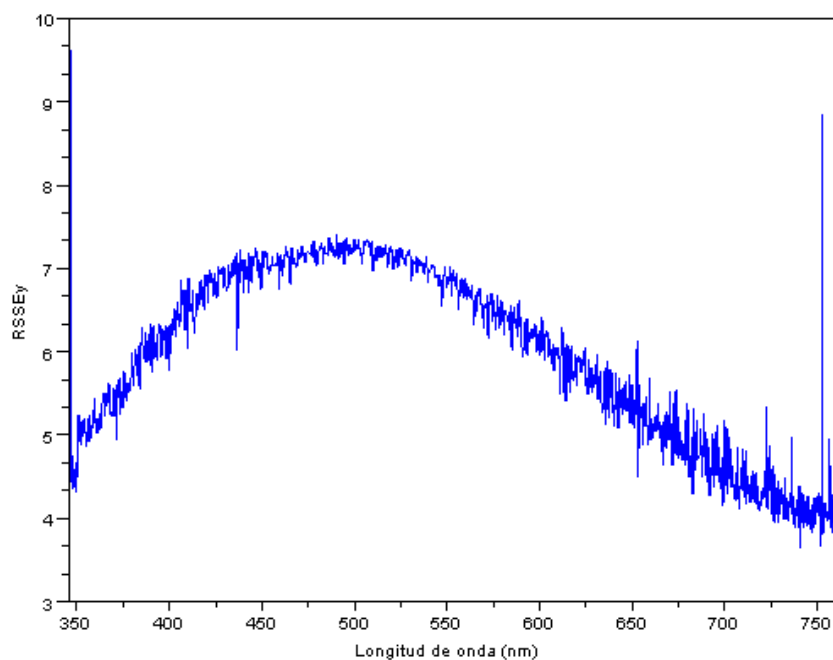


Figura 22. Regresiones locales para asfaltenos

La figura 22 muestra que en la región de longitudes de onda superiores a los 725 nm, y una pequeña zona cerca de los 350 nm, dan los menores errores en las regresiones.

Lo curioso de estos resultados es el parecido entre las zonas de bajo error y las zonas de máxima emisión reportadas en la literatura ¹¹, sobre todo en lo que confiere a los aromáticos y resinas (figura 23).

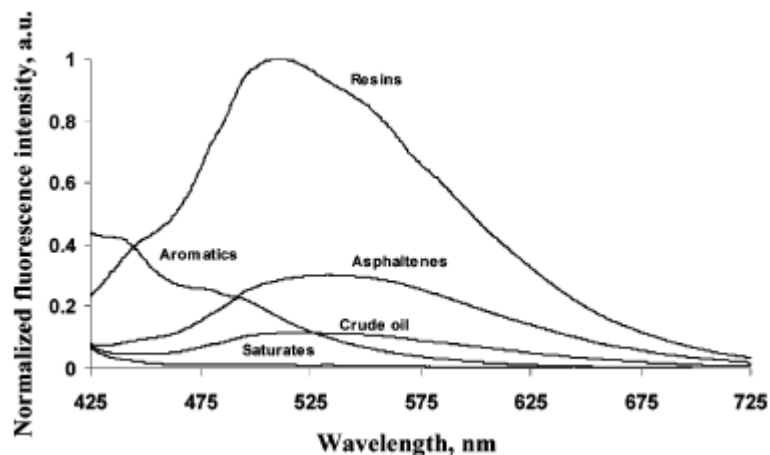


Figura 23. Espectros LIF de las fracciones S.A.R.A. de un crudo Venezolano

Lógicamente la comparación ayuda a entender cómo se comportan las fracciones y no como fiel reproducción común para todos los crudos. Volviendo con esta comparación, los saturados juegan un papel importante coincidiendo con un intervalo de errores mínimo entre las zonas en que también lo hacen los aromáticos y las resinas; por otra parte, una característica muy importante es que las medidas de más bajos errores para los asfaltenos, estén ubicadas donde las otras fracciones presentan errores altos.

El hecho tanto de que la zona de altos errores del asfalteno sean las de errores bajos para las otras fracciones, así como de que la comparación de la figura 23, con la figura 22, no coincidan en la relación bajo error-máxima emisión, indica que los asfaltenos probablemente estén siendo alterados por algún tipo de fenómeno de transferencia de energía.

Otra cosa que se aprovecha para mirar en la figura 23, es que las fracciones por separado fluorescen mucho más que cuando están mezcladas, mostrando la influencia de los fenómenos no radiativos.

3.3.1. Elección de variables principales usando espectros auto-escalados

Esta consiste en la búsqueda manual de características relevantes en cada uno de los espectros, los criterios de selección son máximos, mínimos y puntos de cruzamiento entre pares de espectros auto-escalados (ver figura 24).

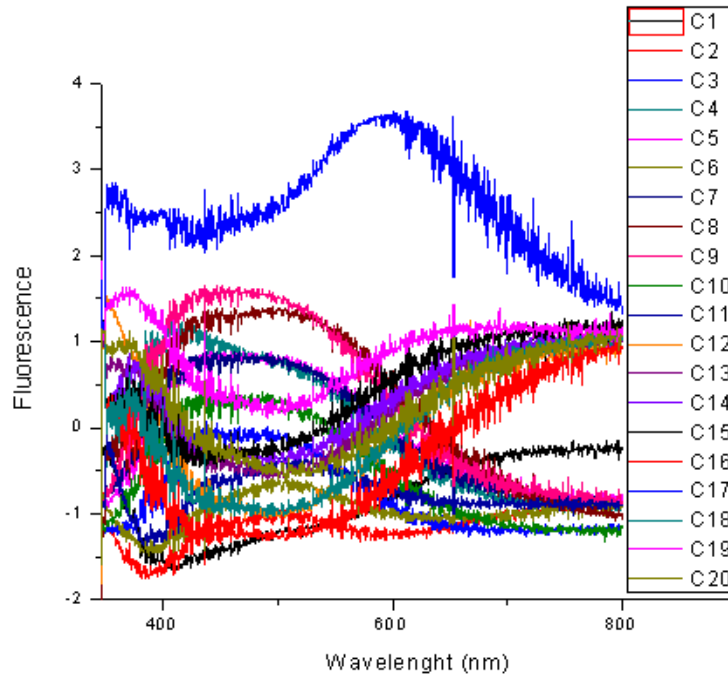


Figura 24. Espectros LIF auto-escalados

Como producto de esta escogencia se obtienen 12 posiciones espectrales, sin embargo, esta tarea no es una opción muy atractiva y por ello, se ha usado la FT con derivación para solucionar esta limitación.

3.3.2. Elección de variables principales en espectros filtrados por FT

Una de las técnicas empleadas en procesamiento de señales es usar la transformada de Fourier (FT) para descomponer una señal en su contenido frecuencial, los componentes de alta frecuencia son extraídos y la señal es reconstruida a partir del contenido frecuencial mas bajo. Teniendo un perfil libre de

ruido de los espectros es mas sencillo usar la primera derivada sobre estos, para lograr extraer las variables principales.

El principal criterio son los ceros del espectro así como sus máximos y mínimos, es así que los espectros filtrados, se derivan y se buscan los máximos y mínimos. Una diferencia entre los espectros es ampliamente notable (ver figuras 25 y 26).

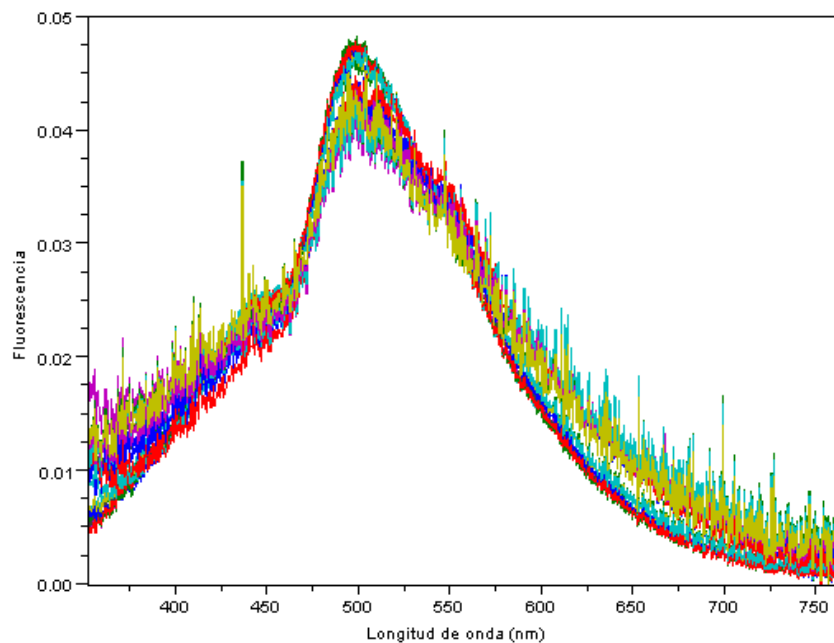


Figura 25. Espectros LIF sin filtrado por FT

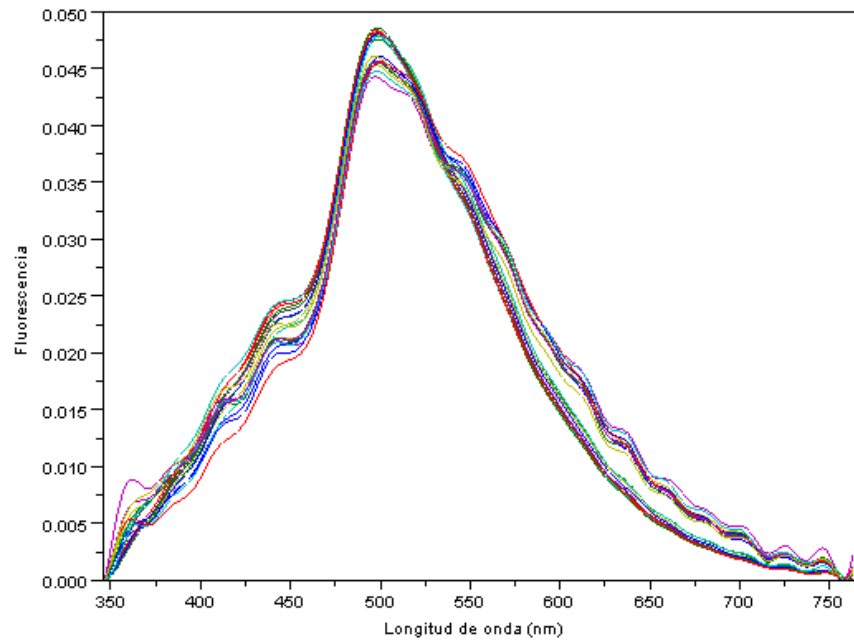


Figura 26. Espectros LIF filtrados por FT

Los espectros de la anterior figura (figura 26) son reconstruidos después de extraer un 99 por ciento del contenido de frecuencias (las frecuencias mas altas).

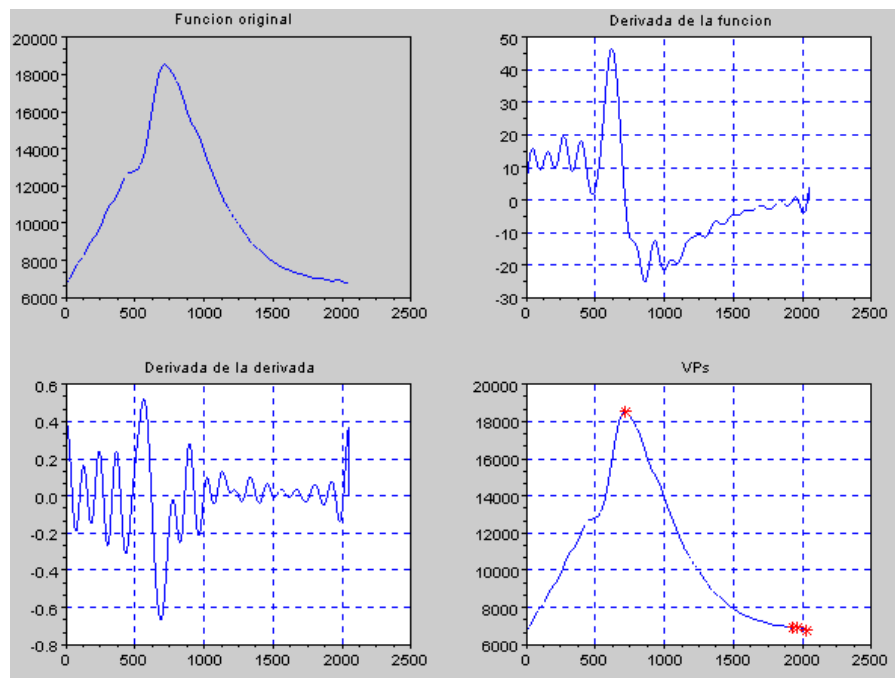


Figura 27. Elección de las variables principales para el crudo C3

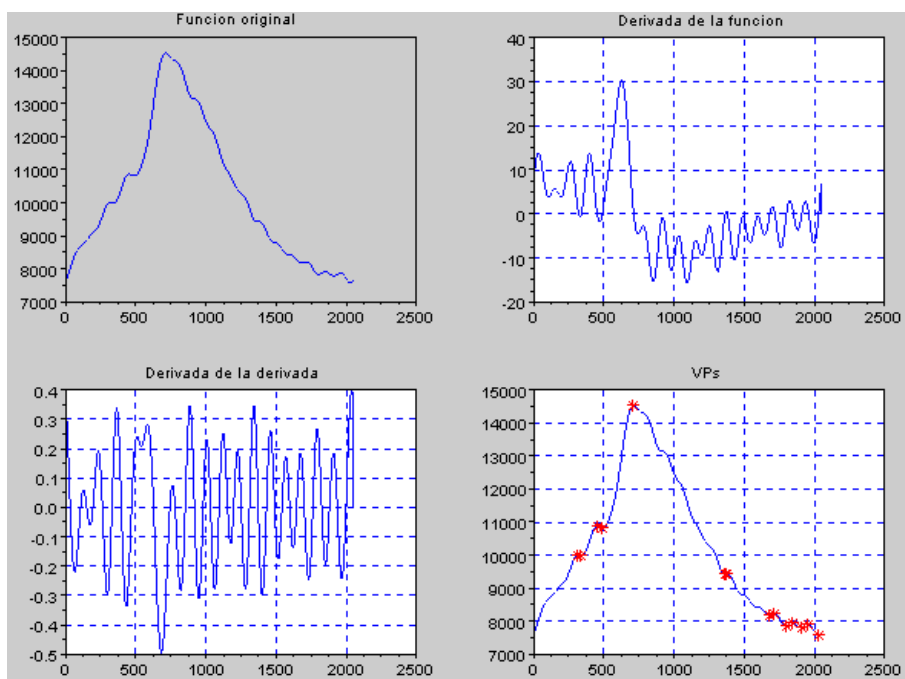


Figura 28. Elección de las variables principales para el crudo C16

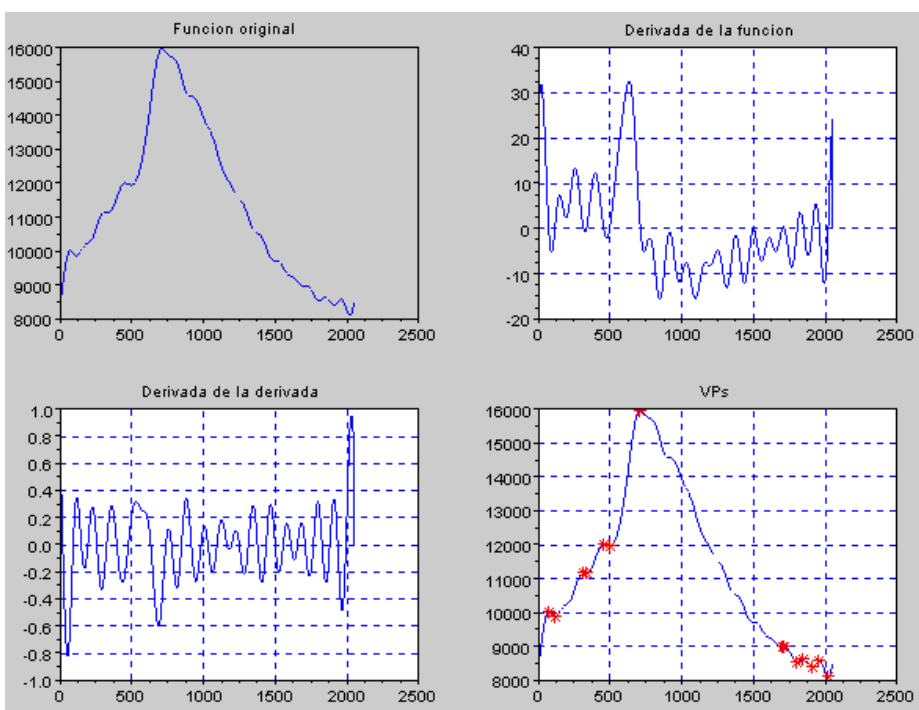


Figura 29. Elección de las variables principales para el crudo C12

Las figuras 27-29, muestran una coincidencia con las figuras 19-22, en que hay preferencias por 3 zonas espectrales principales. A través de la metodología de aplicar FT y derivar, se obtienen números diferentes de variables principales para cada espectro, esto muestra lo diferentes que son los crudos, de manera que, para la construcción del modelo se escogieron las únicas 6 variables que parecen ser comunes para todas las muestras y que por supuesto, dan la mejor estimación de la fracción S.A.R.A. Estas variables están ubicadas en las regiones de bajas energías y en la región del máximo de emisión; como guía de lo dicho, ver la figura (27).

3.3.3. Resultados de la construcción del modelo predictivo

Se encuentra que tanto para la una como para la otra forma de búsqueda y selección de variables, se obtiene un resultado muy similar, pues básicamente el método de FT con derivación devuelve, para este caso, el número e identidad de las variables que mejor predicen S.A.R.A., para datos con una línea base y normalizados a la longitud unidad, simplemente buscando aquellas que sean comunes para todas las muestras; por otra parte, no todas de las 12 variables obtenidas de forma manual desde los espectros auto-escalados, poseen información útil, y por ello, deben ser omitidas para la construcción del modelo, de forma que al final el proceso de descarte lleva a que la identidad del grupo de variables restantes coincida con la identidad de las variables por FT - derivación. Debido a esto, solo se muestra el gráfico de resultados de validación cruzada para las variables encontradas por FT-derivación, por ser un método que cae directamente sobre las variables de interés. Ver figuras 30 hasta la 33.

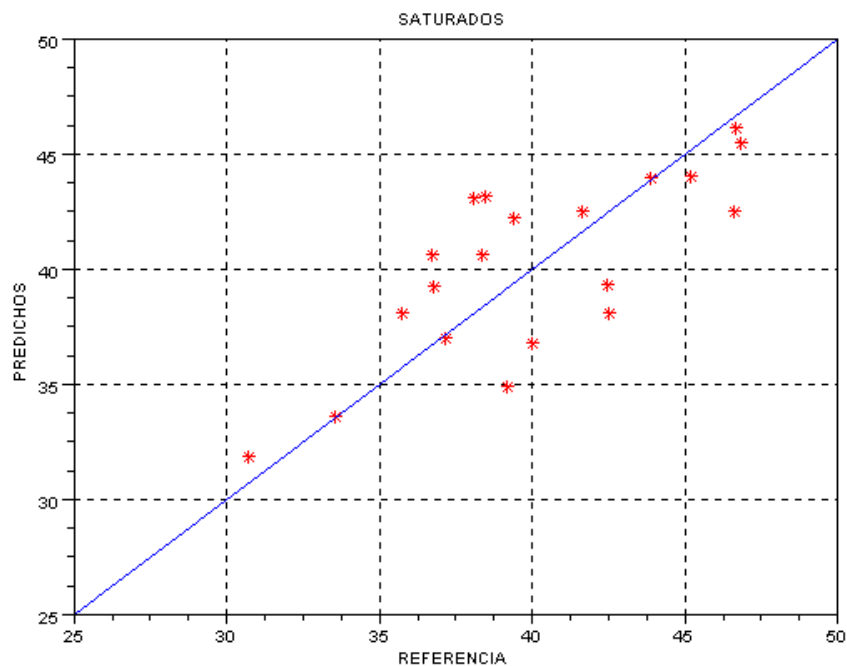


Figura 30. Validación cruzada para variables obtenidas por FT-derivación en saturados.

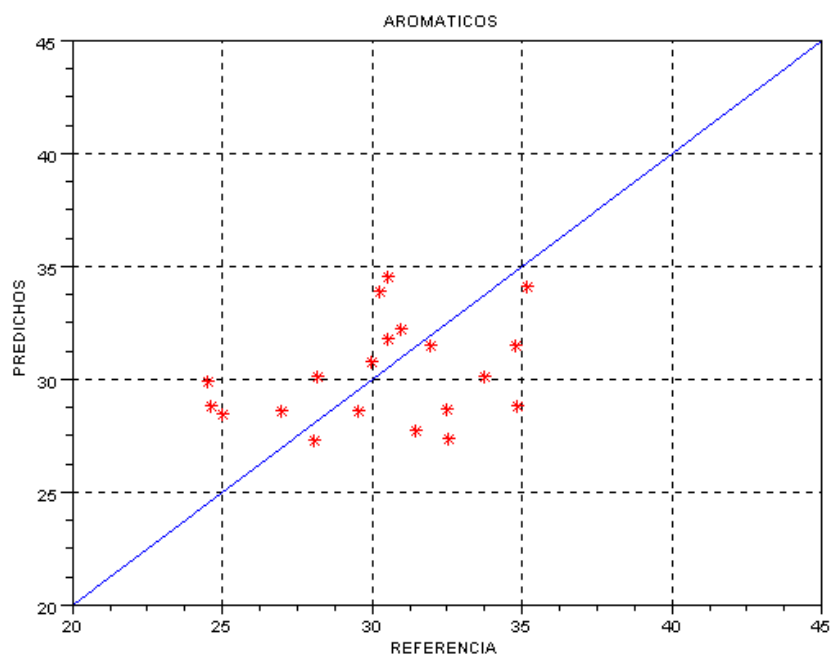


Figura 31. Validación cruzada para variables obtenidas por FT-derivación en aromáticos.

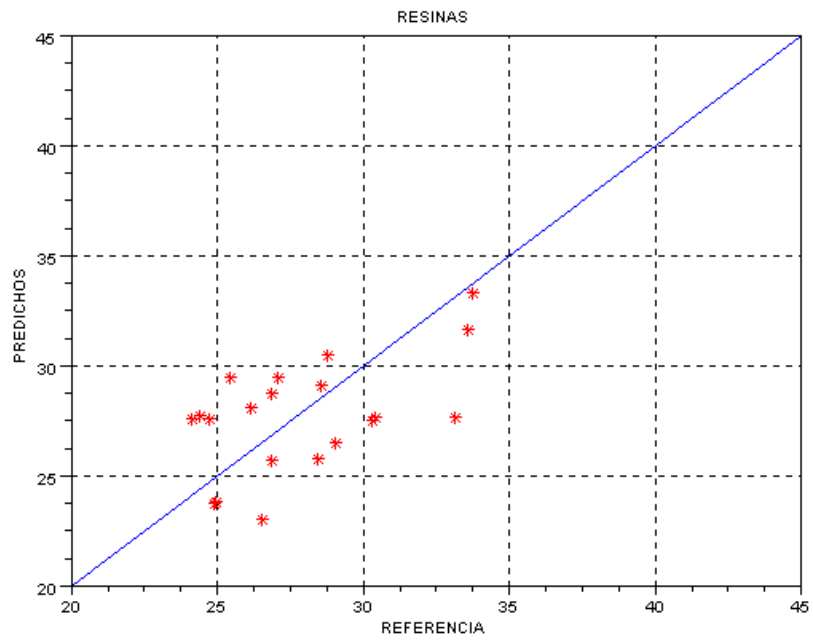


Figura 32. Validación cruzada para variables obtenidas por FT-derivación para resinas.

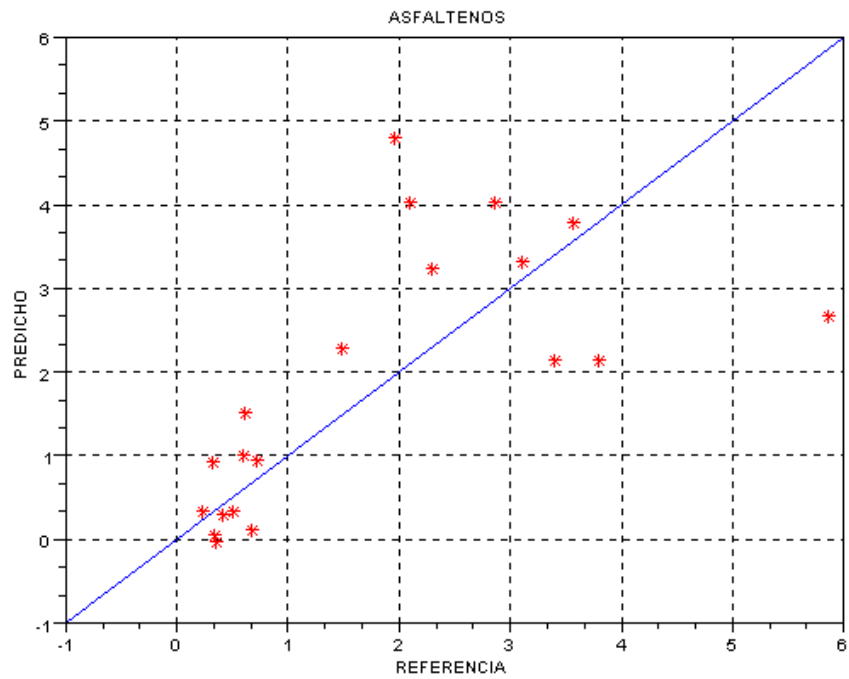


Figura 33. Validación cruzada para variables obtenidas por FT-derivación en asfaltenos.

Las figuras anteriores (figura 30 a figura 33) muestra la relación entre cada una de las fracciones S.A.R.A. de referencia contra los valores predichos por el modelo cuando se usa FT-derivación. La tabla 2, muestra los errores de calibración y de validación cruzada para cada fracción usando las 6 variables del método de FT-derivación.

Tabla 2. Errores para el modelo MLR.

	SATURADOS	AROMATICOS	RESINAS	ASFALTENOS
RMSECV	2,8933992	3,2789896	2,671953	1,2523389
RMSEC	2,0940888	2,3459623	1,929181	0,8158741

Donde RMSEC y RMSECV (ecuación 31 y ecuación 33, respectivamente), son valores aceptables para un sistema tan complejo, en especial teniendo en cuenta que las ya mencionadas variables comunes para todas las muestras tienen pequeños corrimientos de uno a otro espectro (figura 26), que da imprecisión a la posición de la variable y que al incluir todos los valores que dicha variable toma alrededor de todos los espectros, el sistema se vuelve singular o mejor dicho el sistema que se quiere analizar es linealmente dependiente.

3.3.4. Otros resultados

Los resultados obtenidos de PFA, específicamente los tres primeros factores principales, son ordenados de forma que los datos en tres dimensiones queden redefinidos en un sistema de dos dimensiones, una de las medidas es el ángulo en radianes de la recta definida por los valores (scores en PF2, scores en PF3) y (0,0) en el plano de esas dos variables, respecto al eje de scores en PF2; y la otra medida corresponde a PF1. La anterior transformación corresponde a la ecuación 35.

$$\tan^{-1}\left(\frac{PF3}{PF2}\right) = \theta \quad (\text{Ecuación 35})$$

Y como resultado de esta transformación, se produce el gráfico de la figura 34.

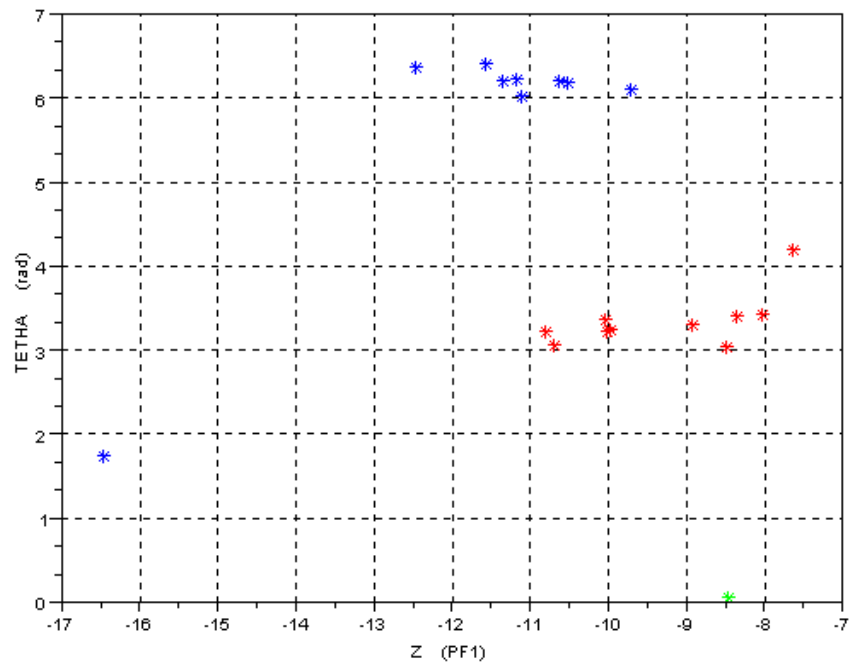


Figura 34. Reordenamiento de datos PFA.

Los resultados obtenidos de PFA están en acuerdo con el reordenamiento mostrado en la figura 34, se quiere mostrar que en los tres primeros PFs persiste la discriminación geográfica vista en la figura 18.

4. CONCLUSIONES

Emplear métodos para la búsqueda de variables principales es una práctica con mas ganancia desde el punto de vista del entendimiento del comportamiento de las fracciones, esto, debido a que es más fácil encontrarle un sentido físico o químico a una variable escogida directamente del espectro, que tratar de encontrarle algún sentido físico o químico a las variables latentes o factores principales que resultan de los métodos de proyección ortogonal como PLS o PFA.

PFA constituye un excelente inicio si de buscar relaciones o comportamientos en las muestras de crudos se trata, al menos desde la perspectiva de los crudos colombianos usando LIF, PFA en alianza con LDA, proveen una herramienta que permite distinguir el origen geográfico de una muestra desconocida a partir de la reducción del espectro de dicho crudo al espacio del grupo de calibración o de construcción, y usar los scores obtenidos en la función discriminante.

El uso de la FT para alisar los espectros para ser derivados y así tener las variables principales, tiene el inconveniente de que el algoritmo trabaja muy eficientemente para espectros con una cantidad de datos expresable como una potencia de 2, lo que implica tener un intervalo espectral tan largo como lo permita el espectrómetro usado. Afortunadamente, el Ocean Optics permite mostrar en este intervalo justo la parte de interés en el espectro.

Notables diferencias en las muestras de petróleo son evidenciadas a lo largo del experimento por varios indicadores: (1) el número de variables que se pudieron encontrar para cada espectro cuando se usó el procedimiento de filtrado FT (diferencia que obligó a tener en cuenta solo las 4 variables comunes a todas las muestras); (2) la superposición de los espectros de fluorescencia (figura 6), muestra diferencias en las intensidades a pesar que los espectros fueron tomados en iguales condiciones experimentales; (3) la tabla 1 muestra notables diferencias en la composición, no solo a nivel general de la población estudiada sino entre

grupos. Es decir, la división de los grupos para construir modelos predictivos para estos crudos colombianos no es conveniente por: (1) no hay una población suficientemente grande de muestras, y (2) dividir significaría reducir el poder predictivo del modelo a situaciones particulares (por ejemplo, solo crudos de una región) distorsionando la característica de lo “multivariable” de los modelos, ya que estos se hacen más fuertes al tener más fuentes de información.

Independientemente de las variables que se escojan parece ser que los asfaltenos no son bien predichos, una causa puede radicar en que los espectros son tomados en un estado sin diluir, según la literatura, estos estados llevan a un dominio de los fenómenos de transferencia de energía sobre la fluorescencia, además estos fenómenos serían más notorios en la fase asfáltica ya que su abundancia relativa es muy pequeña.

Valores del RMS altos no son necesariamente malos, pues este parámetro depende de la magnitud del valor promedio de la variable a la cual se calcula; en la tabla 2, aparentemente el error es menor para los asfaltenos en ambos casos del MLR comparado con el calculado para las otras fracciones, pero lo cierto es que los modelos tienen dificultades para predecir la fracción asfáltica, sin embargo, en caso especial de las variables por derivación, las predicciones para las fracciones saturada, resina, y aromática, son más confiables por lo cual la asfáltica puede ser calculada restando del 100% la suma de las predicciones para las otras fracciones.

La espectroscopia LIF en modo front-face es una excelente opción con responsabilidad y sentido de conservación ecológica y ambiental, ya que evita el uso de solvente tales como heptano, tolueno, etc., además de reducir el riesgo de intoxicaciones, en los operadores que realizan la determinación S.A.R.A., por los métodos establecidos por la ASTM.

La alianza PFA-LDA ha demostrado que puede reconocer la característica de origen geográfico de las muestras de crudos colombianos, que constituye uno de los objetivos más deseados de la quimiometría en análisis de crudos.

5. RECOMENDACIONES

Los resultados del estudio anteriormente presentado muestran el gran potencial de la espectroscopia LIF para la caracterización de la composición SARA de crudos colombianos. Sin embargo para reforzar las habilidades de la sociedad LIF-Quimiometría en este estudio, se recomiendan algunas actividades.

Cuando la composición entre las muestras es tan variable, y a través de la espectroscopia se pueden reconocer características como el origen geográfico es porque de hecho esta última tiene el potencial para reconocer mas variables físico-químicas haciendo posible su predicción; al tener más criterios de relación, la predicción puede en principio mejorar. La introducción de otros parámetros como humedad, contenido de metales pesados, contenido de algún mineral, etc., lograría una mejor relación entre las variables espectroscópicas y las variables de predicción o de propiedades, ya que está demostrado que la fluorescencia es altamente sensible o dependiente de la composición del sistema en análisis.

Los resultados obtenidos por el uso de FT (transformada de Fourier) pueden ser mejorados usando en su lugar WT (Transformada Wavelet), ya que FT supone la existencia del mismo contenido frecuencial en todo el intervalo de tiempo, y WT supone la existencia de estos contenidos frecuenciales con tiempos de aparición diferentes; en otras palabras, WT permitiría hallar los sitios de máxima emisión con más eficiencia, porque tiene la capacidad de reconocer no solo en qué cantidad sucede sino donde sucede la emisión.

Las muestras en este estudio se limitaron a tener en cuenta muestras tomadas en una fecha definida, se sabe que la composición de un pozo varía con la profundidad, al estar siendo extraído el crudo de estos pozos la composición esperada de una fecha a otra cambiará en virtud de los cambios en el nivel del pozo en ese mismo intervalo de tiempo. Se aconseja por ello, renovar constantemente el grupo de muestras de construcción y usar la metodología

ilustrada en este trabajo de investigación o incluir en los modelos el factor de evolución de las propiedades con el tiempo.

Repetir el análisis hecho en este trabajo investigativo, con muchas más muestras permitirá dar más validez al mismo.

6. BIBLIOGRAFÍA

1. FALLA, F.S; C, LARINI; LE ROUX, G.A.C. Characterization of crude petroleum by NIR. Journal of Petroleum Science and Engineering. 2006, 51, 127-137.
2. FUJISAWA, GO; BETANCOURT, S; MULLINS, OLIVER. Hydrocarbon Compositional Gradient Revealed by In-Situ Optical Spectroscopy. SPE Reservoir Evaluation & Engineering. 2008, 233-237.
3. SPEIGHT, J. G. The chemistry and technology of petroleum. Taylor and Francis Group, LLC. Cuarta edición. 2007.
4. SIMANZHENKOV, V. Crude Oil Chemistry. Marcel Dekker, Inc. USA, 2003.
5. BEHRENBRUCH, P; DEDIGAMA, T. Classification and characterisation of crude oils based on distillation properties. Journal of Petroleum Science and Engineering. 2007, 57, 166-180.
6. LAKOWICZ, J. R. Principles of fluorescence spectroscopy. Editorial Springer. Tercera edición. New York. 2006. 960p.
7. PARSON, WILLIAM. Modern Optical Spectroscopy. Springer. USA, 2007.
8. SKOOG, DOUGLAS. Principios de Análisis Instrumental. Quinta Edición. Mc Graw Hill. España, 2001.
9. GIRI, RAJESH. Fluorescence quenching of coumarins by halide ions. Spectrochimica Acta Part A. 2006, 60, 757-763.
10. TKACHENKO, NIKOLAI. Optical Spectroscopy: Methods and Instrumentations. Primera Edición. Elsevier. Netherland, 2006.
11. RIVEROS, LISETH; JAIMES, BLANCA; RANAUDO, MARIA; CASTILLO JIMMY; CHIRINOS JOSE. Determination Asphaltene and Resin Content in Venezuelan Crude Oils by Using Fluorescence Spectroscopy and Partial Least Squares Regression. Energy & Fuels. 2006, 20, 227-230.

12. PRADIER, B; LARGEAU, C; DERENN, S; MARTINEZ, L; BERTRAND, P. Chemicals Basis of Fluorescence Alteration of oils and Kerogens. *Org. Geochem.* 1990, 16, 451-460.
13. WANG, X; MULLINS, O.C. Fluorescence Lifetimes studies of crude oils. *Applied Spectroscopy.* 1994, 48, 977-984.
14. CAMAGNI, P; COLOMBO, A; KOECHLER, C. Fluorescence response of minerals oils: Spectra yield vs absorption and decay time. *Appl. Opt.* 1991, 30, 26-35.
15. Sociedad de Quimiometría y Cualimetría de España. (En Línea). <http://www.ub.edu/seqyc/> [citado el 1 diciembre de 2009].
16. RASMUS, BRO. Multivariate Calibration: What is in chemometrics for the analytical chemist?. *Analytica Chimica Acta.* 2003, 500, 185-194.
17. GELADI, PAUL. Chemometrics in Spectroscopy. Part 1. Classical Chemometrics. *Spectrochimica Acta Part B.* 2003, 58, 767-782.
18. MUTHILAC, LUCIA. Mining in chemometrics. *Analytica Chimica Acta.* 2008, 612, 1-18.
19. WOLD, SVANTE. Chemometrics, present and future success. *Chemometrics and Intelligent laboratory Systems.* 1998, 44, 3-14.
20. LAKOWICZ, JOSEPH. Principles of Fluorescence Spectroscopy. Tercera Edición. Springer. USA, 2006.
21. SAVITSKY, A; GOLAY, MARCEL. Smoothing and Differentiation of data by Simplified least Squares Procedures. *Analytical Chemistry.* 1964, 36, 1627-1639.
22. HOWERTON, SAMUEL; GOODPASTER, V; Mc GUFFIN, VICTORIA. Characterization of Polycyclic aromatic hydrocarbons in environmental samples by selective fluorescence quenching. *Analytica Chimica Acta.* 2002, 459, 61-73.

23. RYDER, A. G. Quantitative Analysis of crude oils by fluorescence lifetime and steady state measurements using 380 nm excitation. *Applied Spectroscopy*. 2002, 56, 107-116.
24. SAVITSKY, S; GOLAY. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*. 1964, 36, 1627-1639.
25. OPPENHEIM, ALAN. Señales y Sistemas. Segunda Edición. Prentice Hall. USA, 2007.
26. CHAU, FOO-TIM; LIANG, YI-ZANG. Chemometrics from basics to wavelet transform. Wiley Interscience. USA, 2004.
27. BURDEN, RICHARD; FAIRES, J. Analisis Numérico. Séptima Edición. Editorial MATH. 2000.
28. MALINOWSKI, EDMUND. Factor Analysis in Chemistry. Tercera Edición. Wiley Interscience. USA, 2002.
29. HOPKE, PHILIP. The evolution of chemometrics. *Analytica Chimica Acta*. 2003, 500, 365-377.
30. KETTANEH, NOUNA; BERGLUND, ANDERS; WOLD SVANTE. PCA and PLS with very large data sets. *Computational Statistics & Data Analysis*. 2005, 48, 69-85.
31. MILLER, JAMES; MILLER, JANE. Estadística para Química Analítica. Prentice Hall. Cuarta Edición. Madrid, 2002.
32. HOLLAS, MICHAEL. Modern Spectroscopy. Cuarta edición. Wiley Interscience. England, 2004.
33. FEIPING, NIE. Semi-supervised orthogonal discriminant analysis via label propagation. *Pattern Recognition*. 2009, 42, 2615-2623.
34. LIANG , ZHIZHENG. Kernel direct discriminant analysis and its theoretical foundation. *Pattern Recognition*. 2005, 38, 445-447.

35. NOORGARD, L; SAUDLAND, J; NIELSEN, J; MUNCK, L; ENGELSEN, S. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy*. 2000, 54, 413-419.
36. JIANG, JIAN-HUI; OZAKI, YUKIHIRO. Wavelength Interval Selection in Multicomponent Spectral Analysis by Moving Window Partial Least-Squares Regression with Applications to Mid-Infrared and Near-Infrared Spectroscopic Data. *Analytical Chemistry*. 2002, 74, 3555-3565.
37. KAROUI, R; DUFOUR, E. Front face fluorescence spectroscopy coupled with chemometric tools for monitoring the oxidation of semi-hard cheeses throughout ripening. *Food Chemistry*. 2007, 101, 1305-1314.
38. DUFOUR, E; FRENCIA, JEAN. P. Development of rapid method base don front-face fluorescence spectroscopy for the monitoring of fish freshness. *Food Research International*. 2003, 36, 415-423.
39. KANOUI, ROMDHANE; SCHOONHEYDT, R. *Analytica Chimica Acta*. 2007, 582, 83-91.
40. GATELLIER, P; GOMEZ, S; GIGAUD, V. Use of fluorescence front face technique for measurement of lipid oxidation during refrigerated storage of chicken meat. *Meat Science*. 2007, 76, 543-547.
41. LE MOIGNE, M; DUFOUR, E. Front face spectroscopy and visible spectroscopy coupled with chemometrics have the potential to characterise ripening of Cabernet Franc grapes. *Analytica Chimica Acta*. 2008, 621, 8-18.
42. RIZKALLAH, JAD; MORALES, F; FOGLIANO, V. Front face fluorescence and multiway analysis for process control and NFC prediction in industrially processed cookies. *Chemometrics and Intelligent Laboratory Systems*. 2008, 93, 90-107.

43. HAMMAMI, M; ROUISSI, H; SALH, N. Fluorescence spectroscopy coupled with factorial discriminant analysis technique to identify sheep milk from different feeding systems. *Food Chemistry*. 2010, xx, xxx-xxx.
44. DIEZ, R; ORTIZ, M; SARABIA, L. Potential of front face fluorescence associated to PLS regression to predict nutritional parameters in heat treated infant formula models. *Analytica Chimica Acta*. 2008, 606, 151-158.
45. PEREIRA, RITA; SKROBOT, VINICIUS; CASTRO, E; FORTES, I; PASA, V. Determination of Gasoline Adulteration by Principal Components Analysis-Linear Discriminant Analysis Applied to FTIR Spectra. *Energy & Fuels*. 2006, 20, 1097-1102.
46. CHEN, ZENG-PING; LIU, ZHUO; CAO, YU-ZHEN; YU, RU-QIN. Efficient way to estimate the optimum number of factors for trilinear decomposition. *Analytica Chimica Acta*. 2001, 444, 295-307.
47. SARMA, A; RYDER, A. Comparison of the fluorescence behavior of a biocrude oil and crude petroleum oils. *Energy & Fuels*. 2006, 20, 783-785.
48. FERRER, R; BELTRAN, J; GUITERAS, J. Multivariate calibration applied to synchronous fluorescence spectrometry. Simultaneous determination of polycyclic aromatic hydrocarbons in water samples. *Talanta*, 1998, 45, 1073-1080.
49. SRIVIDYA, N; RAMAMURTHY, P; RAMAKRISHNAN, V. Solvent effects on the absorption and fluorescence spectra of some acridinedione dyes: determination of ground and excited state dipole moments. *Spectrochimica Acta Part A*. 1997, 53, 1743-1753.
50. MA, XIAOLIANG; SAKANISHI, K; ISODA, T; NAGAO, S; MOCHIDA, I. Structural characteristics and removal of visible-fluorescence species in hydrodesulfurized diesel oil. *Energy & Fuels*. 1996, 10, 91-96.

51. FEDOROV, V. Studying organic species in water by laser fluorescence spectroscopy with a source of excitation in mid-UV range (266 nm). *Water Resources*. 2005, 32, 549-554.
52. MORGAN, T; MILLAN, M; BEHROUZI, M; HEROD, A; KANDIYOTI, R. On the limitations of UV-fluorescence spectroscopy in the detection of high-mass hydrocarbon molecules. *Energy & Fuels*. 2005, 19, 164-169.
53. SCHNEIDER, M; ANDREWS, A; MULLINS, O. Asphaltenes molecular size by fluorescence correlation spectroscopy. *Energy & Fuels*. 2007, 21, 2875-2882.
54. CHI, ZHEHUAN; CULLUM, BRIAN; STOKES, D; MOBLEY, J; MILLER, G; HAJALIGOL, M; VO-DINH, T. Laser-induced fluorescence studies of polycyclic aromatic hydrocarbons (PAH) vapors at high temperatures. *Spectrochimica Acta Part A*. 2001, 57, 1377-1384.
55. DIXON, SARAH; BRERETON, RICHARD. Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, Linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure. *Chemometrics and Intelligent Systems*. 2009, 95, 1-17.
56. WOLD, SVANTE; SJOSTROM, M; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Systems*. 2001, 58, 109-130.
57. GELADI, P; KOWALSKI, B. Partial least-squares regression: A tutorial. *Analytica Chimica Acta*. 1986, 185, 1-17.
58. LEE, C. K; KO, E. J; KIM, K. W; KIM, Y. J. Partial least square regression method for the detection of polycyclic aromatic hydrocarbons in the soil environment using laser-induced fluorescence spectroscopy. *Water, Air, and Soil Pollution*. 2004, 158, 261-275.
59. HONORATO, F; NETO, B; PIMENTEL, M; STRAGEVITCH, L. Using principal component analysis to find the best calibration settings for simultaneous

spectroscopic determination of several gasoline properties. Fuel. 2008, 87, 3706-3709.

60. Web Scilab. (En Línea). <http://www.scilab.org> [citado el 10 marzo 2010].

61. GRUPO DE QUIMIOMETRÍA Y CUALIMETRÍA DE TARRAGONA. Quimiometría: Una disciplina útil para el análisis químico. (En línea). <http://www.quimica.urv.es/quimio> [citado el 24 enero 2010].

7. ANEXOS

7.1. ALGORTIMOS SCILAB

Algoritmo para PFA

```
//-----PFA(svd)-----//
l1=list('Ubicación de la variable',2,['Consola','Disco']);
Pregun=x_choices('',list(l1));
if Pregun(1)==2 then
    rep=uigetfile(['*.txt'; '*.sce'; '*.xls*']);
    cod=str2code(rep);
    clear rep;
    [fd,SST,Sheetnames,Sheetpos]=xls_open('+code2str(cod)+')
    //lee la primera hoja de datos o espectroscopica
    [data,TextInd] = xls_read(fd,Sheetpos(1))
    //cierra el flujo de la hoja de calculo
    mclose(fd)
    // se ha leído la matriz de datos espectroscopicos
    clear cod;
end,
if Pregun(1)==1 then
    data=data;
end,
//-----//
stacksize(59000000);
l1=list('Espectros a lo largo de',1,['filas','columnas']);
Pregunta=x_choices('',list(l1)); //pregunta como estan ordenados los datos, si los espectros estan
como filas o como columnas
if Pregunta(1)==1 then
    data=data';
end;
[f c]=size(data);
disp('Ya están leídas las matrices de datos!!');
//-----//
// 4 formas de pre-procesamiento de los datos:
n = x_choose(['1. Covarianza en el origen';'2. Covarianza en el centro';'3. Correlacion en el
origen';'4. Correlacion en el centro';'5. Autoescalado (para LIF)'],['Tipo de Pre-
procesamiento'],'Salir');
//-----//
select n,
case 1 then
    M=data',
// la matriz de covarianzas en origen
```

```

//-----//
case 2 then
    [f c]=size(data);
    prom=mean(data, 'c');
    // Promedio a cada longitud, o sea, por filas
    // Desviacion estandar a cada longitud, o sea, por filas
    for j=1:c,
        data(:,j)=data(:,j)-prom;
    end;
    clear prom;
    stacksize('max');
    M=data',
    // Aqui se ha calculado la matriz de covarianzas centrada
    //-----//
case 3 then
    [f c]=size(data);
    for i=1:f,
        escal(i,1)=norm(data(i,:));
    end;
    for j=1:c,
        data(:,j)=data(:,j)./escal;
    end;
    clear escal;
    stacksize('max');
    M=data',
    // calculado la matriz de correlaciones en origen
    //-----//
case 4 then
    [f c]=size(data);
    prom=mean(data, 'c');
    // Promedio a cada longitud, o sea, por filas
    for i=1:f,
        escal(i,1)=norm(data(i,:)-prom(i));
    end;
    for j=1:c,
        data(:,j)=(data(:,j)-prom)./escal;
    end;
    clear escal; clear prom;
    stacksize('max');
    M=data',
    // Matriz de correlaciones centradas
    //-----//
case 5 then
    [f c]=size(data);
    prom=mean(data, 'c');
    // Promedio a cada longitud, o sea, por filas

```

```

for i=1:f,
    escal(i,1)=sqrt((sum((data(i,:)-prom(i))**2))/(c-1));
end;
for j=1:c,
    data(:,j)=(data(:,j)-prom)./escal;
end;
clear escal; clear prom;
stacksize('max');
M=data',
// Matriz autoescalada
//-----//
end;
disp('Termina pre-tratamiento!!');
// Termina pre-tratamiento de datos
//-----//
stacksize('max');
[U,S,V] = svd(M);
[f c]=size(S);
clear V;
score=U*S; clear U; clear S;
//se han calculado los eigenvectores
//-----//
g = x_choose(['2-D';'3-D'],['Tipo de Grafico'],'Salir');
if g==1 then
    txt=['Eje x';'Eje y'];
    sig=x_mdialog('Factores a graficar',txt,['1';'2'])
    f1=evstr(sig(1))
    f2=evstr(sig(2))
    if f1>=0 & f1<=c then
        if f2>=0 & f2<=c then
            messagebox('Valores escogidos >=0 y <='+string(c)+'');
            txt=['Eje x';'Eje y'];
            sig=x_mdialog('Factores a graficar',txt,['1';'2'])
            f1=evstr(sig(1))
            f2=evstr(sig(2))
        end,
    end,
    plot(score(:,f1),score(:,f2),'*r');
    xgrid;
elseif g==2 then
    txt=['Eje x';'Eje y';'Eje z'];
    sig=x_mdialog('Factores a graficar',txt,['1';'2';'3'])
    f1=evstr(sig(1))
    f2=evstr(sig(2))
    f3=evstr(sig(3))
    if f1<0 or f1>c then

```

```

if f2<0 or f2>c then
  if f3<0 or f3>c then
    messagebox('Valores escogidos >=0 y <='+string(c)+'');
    txt=['Eje x';'Eje y';'Eje z'];
    sig=x_mdialog('Factores a graficar',txt,['1';'2';'3'])
    f1=evstr(sig(1))
    f2=evstr(sig(2))
    f3=evstr(sig(3))
  end,
end,
end,
plot3d(score(:,f1),score(:,f2),score(:,f3));
end,
//Termina el programa de PFA (Análisis de Factores Principales)

```

Algoritmo para LDA

```

//-----LDA(PCA(svd))-----//
//los datos de entrada son como vectores columnas en una matriz
//llamada z
//para 2 grupos
//los datos deben estar ordenados en z como:
//espectros por fila y variables discriminantes por columna

txt=['Desde';'Hasta'];
sig=x_mdialog('Grupo 1',txt,['2';'11'])
lim11=evstr(sig(1))
lim12=evstr(sig(2))
//lectura limites del grupo 1
txt=['Desde';'Hasta'];
sig=x_mdialog('Grupo 2',txt,['12';'20'])
lim21=evstr(sig(1))
lim22=evstr(sig(2))
//-----//
[f c]=size(z);
cent=mean(z,'r');
for i=1:f,
  X(i,:)=(z(i,:))-(cent);
end,
//centrado de las variables discriminantes originales
//-----//
//lectura limites del grupo 2
[f c]=size(X); g1=0; g2=0;
for i=1:f,
  if i>=lim11 & i<=lim12 then
    g1=g1+1;

```

```

    G1(g1,:)=X(i,:);
end,
if i>=lim21 & i<=lim22 then
    g2=g2+1;
    G2(g2,:)=X(i,:);
end,
end,
//separacion de los 2 grupos
//-----//
[n1 c]=size(G1); [n2 c]=size(G2);
a=[n1 n2]; clear n1; clear n2;
Dg=diag(a);
xm1=mean(G1,'r'); xm2=mean(G2,'r'); // vector de medias
Xm=[xm1;xm2]; //matriz con las medias de los grupos
Qh=Xm'*Dg*Xm; //sumas de cuadrados entre grupos
Qt=X'*X; //suma de cuadrados totales
Qe=Qt-Qh; //suma de cuadrados dentro de los grupos
St=(1/(f-1))*Qt; //matriz de covarianzas total
g=2; //dos grupos (g = # de grupos de muestras)
Sh=(1/(g-1))*Qh; //matriz de covarianzas entre grupos
S=(1/(f-g))*Qe; //matriz de covarianzas dentro de los grupos
//-----//
//-----//
coefs=((xm1-xm2)*(inv(S)))';
cort=(1/2)*((xm1+xm2)*(coefs));
//parametros funcion discriminante (W):
// W = (x)*(coefs) - (cort)
//-----//
l1=list('Datos para probar en disco?',1,['Si','No']);
Pregunta=x_choices('',list(l1)); //pregunta como estan ordenados los datos, si los espectros estan
como filas o como columnas
if Pregunta(1)==1 then
    rep=uigetfile(['*.txt'; '*.sce'; '*.xls*']);
    cod=str2code(rep); clear rep;
    [fd,SST,Sheetnames,Sheetpos]=xls_open(""+code2str(cod)+"")
    //lee la primera hoja de datos o espectroscopica
    [data,TextInd] = xls_read(fd,Sheetpos(1))
    //cierra el flujo de la hoja de calculo
    mclose(fd)
    // se ha leido la matriz de datos espectroscopicos
    clear cod;
    //-----//
    l1=list('Muestras por fila?',1,['Si','No']);
    Pregunta1=x_choices('',list(l1));
    if Pregunta1(1)==2 then
        data=data';

```

```

end,
[f c]=size(data);
for i=1:f,
    x(i,:)=(data(i,:))-(cent);
end,
clear data;
g1=0; g2=0; g=0;
for i=1:f,
    w=((x(i,:))*(coefs))-(cort);
    if w>0 then
        g1=g1+1;
        pos1(g1,1)=i;
    elseif w<0 then
        g2=g2+1;
        pos2(g2,1)=i;
    else
        g=g+1;
        pos(g,1)=i;
    end,
end,
//-----//
//muestra los resultados de la discriminacion
disp('Muestras grupo1 ='); disp(pos1);
disp('Muestras grupo2 ='); disp(pos2);
if g~=0 then
    disp('Muestras con problemas de asignacion ='); disp(pos);
elseif g==0 then
    disp('Sin problemas de asignacion!!');
end,
end;

```

Algoritmo para filtro FT para espectros LIF

```

//-----Filtro FT para espectros LIF-----//
l1=list('Ubicación de la variable',2,['Consola','Disco']);
Pregun=x_choices('',list(l1));
if Pregun(1)==2 then
    rep=uigetfile(['*.txt'; '*.sce'; '*.xls*']);
    cod=str2code(rep);
    clear rep;
    [fd,SST,Sheetnames,Sheetpos]=xls_open('+code2str(cod)+')
    //lee la primera hoja de datos o espectroscopica
    [data,TextInd] = xls_read(fd,Sheetpos(1))
    //cierra el flujo de la hoja de calculo
    [Y,TextInd] = xls_read(fd,Sheetpos(2))
    mclose(fd)

```

```

// se ha leído la matriz de datos espectroscopicos
clear cod;
end,
if Pregun(1)==1 then
    data=data;
end,
//-----//
l1=list('Espectros a lo largo de',2,['filas','columnas']);
Pregunta=x_choices('',list(l1));
if Pregunta(1)==1 then
    data=data';
end;
// Se ordenan los datos
// Espectros como vectores columna
stacksize(59000000); // Aumenta la memoria disponible
[f c]=size(data);
lo=log2(f); // La condición de potencias de 2, para usar fft()
lo1=round(lo); // Entero menor más cercano
lo1=2**(lo1); // nuevo tamaño de vector para condición de potencia
espectros=zeros(lo1,c); // re-dimensionamiento de la matriz espectral
// los espacios faltantes se llenan con ceros
if lo1>=f then
    for i=1:f,
        espectros(i,:)=data(i,:);
    end,
clear data; // libera la memoria de la matriz completa
x1=[0.001:.001:(lo1*0.001)]; // variable independiente con
// aspecto de continua
labels=[" Pasa-bajos, Nivel: 1%-100% "];
[ok,corte]=getvalue("Nivel(%)",labels,...
list("vec",1),["100"]);
l1=list('Quisiste decir... ('+string(corte)+)',1,['Si','No']);
Pregunta3=x_choices('',list(l1));
if Pregunta3(1)==1 then
    [v e]=size(espectros);
    corte=round((v/100)*corte);
    for i=1:c,
        espectro=espectros(:,i);
        y=fft(espectro); // Transformada rápida de Fourier
        [a b]=size(y);
        for j=corte:a,
            y(j)=0;
        end,
        xx1=mtlb_ifft(y); // transformada inversa de Fourier
        espectros(:,i)=xx1;
    end,
end,

```

```

elseif Pregunta3(1)==2 then
    messagebox('Solo numeros enteros!!');
    disp('Mision abortada!!');
    abort;
end,
elseif lo1<f then
    for i=1:lo1,
        espectros(i,:)=data(i,:);
    end,
    //clear data; // libera la memoria de la matriz completa
    x1=[0.001:.001:(lo1*0.001)];
    labels=[" Pasa-bajos, Nivel: 1%-100% "];
    [ok,corte]=getvalue("Nivel(%)",labels,...
    list("vec",1),["100"]); // se especifica el nivel de filtrado
    l1=list('Quisiste decir... ('+string(corte)+)',1,['Si','No']);
    Pregunta3=x_choices(",list(l1)); //verifica el valor
    if Pregunta3(1)==1 then
        [v e]=size(espectros);
        corte=round((v/100)*corte); //relacion porcentual
        for i=1:c,
            espectro=espectros(:,i);
            y=fft(espectro); // Transformada rapida de Fourier
            [a b]=size(y);
            for j=corte+1:a,
                y(j)=0;
            end,
            xx1=mtlb_ifft(y); // tranformada inversa de Fourier
            espectros(:,i)=xx1;
        end,
    elseif Pregunta3(1)==2 then
        messagebox('Solo numeros enteros!!');
        disp('Mision abortada!!');
        abort;
    end,
    //plot(x1,espectros);
    //title('Espectros Alisados con FT');
    //xgrid;
end,
//-----//
l1=list('Deseas trazar linea base??',1,['Si','No']);
Pregunta3=x_choices(",list(l1));
if Pregunta3(1)==1 then //se valida en caso de respuesta afirmativa
    [a b]=size(espectros);
    for i=1:b,
        zurda=real(espectros(1:200,i)); diest=real(espectros(1948:2048,i));
        [a1 b1]=min(zurda); [a2 b2]=min(diest);
    end,
end,

```

```

m=(a2-a1)/((b2+1848)-b1);
// y-y1=m(x-x1)-----y=mx+(y1-mx1)...relacion usada
inte=a1-m*b1; //corte con la ordenada del espectro
function[y]=lineb(j) //funcion sustraccion de linea base
    y=(m*j)+inte
endfunction
for j=1:a,
    [y]=lineb(j);
    espectros(j,i)=(espectros(j,i))-(y);
end,
end,
//figure()
//plot(x1,espectros);
//title('Espectros con linea base lineal');
//xgrid;
elseif Pregunta3(1)==2 then
    messagebox('Pues se lo pierde!!');
    abort;
end,
//Termina la linea base
//-----//
l1=list('Deseas normalizar a longitud unidad??',1,['Si','No']);
Pregunta3=x_choices(' ',list(l1));
if Pregunta3(1)==1 then
    espectros=real(espectros);
    for i=1:b,
        esc(1,i)=norm(espectros(:,i));
    end,
    for j=1:a,
        espectros(j,:)=(espectros(j,:))./esc;
    end,
    //figure()
    //plot(x1,espectros);
    //title('Espectros Normalizados a longitud unidad');
    //xgrid;
elseif Pregunta3(1)==2 then
    messagebox('Pues Pailas!!');
    abort;
end,
//termina normalizacion a longitud unidad
//-----//

```

Algoritmo para búsqueda de VPs (variables principales)

```

//-----Search VP's-----//
l1=list('Ubicación de la variable',2,['Consola','Disco']);
Pregun=x_choices('',list(l1));
if Pregun(1)==2 then
    rep=uigetfile(['*.txt'; '*.sce'; '*.xls*']);
    cod=str2code(rep);
    clear rep;
    [fd,SST,Sheetnames,Sheetpos]=xls_open('+code2str(cod)+')
    //lee la primera hoja de datos o espectroscopica
    [data,TextInd] = xls_read(fd,Sheetpos(1))
    //cierra el flujo de la hoja de calculo
    mclose(fd)
    // se ha leído la matriz de datos espectroscopicos
    clear cod;
end,
if Pregun(1)==1 then
    //data=data; //cuando los datos vienen de consola, se active esta opcion
    data=real(espectros); //cuando los datos vienen de Fourier matrices
end,
//-----//
stacksize(59000000);
l1=list('Espectros a lo largo de',2,['filas','columnas']);
l2=list('Valores referencia a lo largo de',1,['filas','columnas']);
Pregunta=x_choices('',list(l1,l2)); //pregunta como estan ordenados los datos, si los espectros
estan como filas o como columnas
if Pregunta(1)==2 then
    data=data';
end;
if Pregunta(2)==2 then
    Y=Y';
end;
//matriz espectroscopica con espectros por filas
//matriz de propiedades con valores de referencia por filas
//se lee la matriz de propiedades
//-----//
[f c]=size(data); [f1 c1]=size(Y);
if f==f1 then
    labels=["Hay (" +string(c1)+") Propiedades"];
    [ok,n]=getvalue("Cual de ellas?",labels,...
    list("vec",1),["1"]);
    y=Y(:,n);
    for i=1:c,
        x=data(:,i);
        if det(x'*x)==0 then
            err(i,1)=1000;
        end,

```

```

if det(x'*x)~=0 then
    m1=inv(x'*x);
    m2=m1*x';
    coef=m2*y;
    dif=y-(x*coef);
    err(i,1)=sqrt(sum(dif*dif));
end,
end,
end,
//-----//

```

Algoritmo para hacer la validación cruzada de MLR

```

//-----MLR--validacion cruzada-----//
l1=list('Ubicación de la variable',2,['Consola','Disco']);
Pregun=x_choices('',list(l1));
if Pregun(1)==2 then
    rep=uigetfile(['*.txt'; '*.sce'; '*.xls*']);
    cod=str2code(rep);
    clear rep;
    [fd,SST,Sheetnames,Sheetpos]=xls_open(""+code2str(cod)+"")
    //lee la primera hoja de datos o espectroscopica
    [Y,TextInd] = xls_read(fd,Sheetpos(1))
    //cierra el flujo de la hoja de calculo
    [pv,TextInd] = xls_read(fd,Sheetpos(2))
    //cierra el flujo de la hoja de calculo
    mclose(fd)
    //se ha leído la matriz de datos espectroscopicos
    clear cod;
end,
if Pregun(1)==1 then
    //data=data;
    data=real(espectros); //cuando los datos vienen de Fourier matrices
end,
//-----//
l1=list('Espectros a lo largo de',2,['filas','columnas']);
Pregunta=x_choices('',list(l1));
if Pregunta(1)==2 then
    data=data';
end;
//para acomodarlos como vectores fila
//-----//
//comienza la regresion con validacion cruzada
[e1 e2]=size(pv);
j=pv(1);

```

```

X=data(:,j);
for t=2:e1,
    j=pv(t);
    X=[X data(:,j)]; //vectores de construccion con variable escogida
end,
[a b]=size(X);
for k=1:a,
    if k==1 then
        Dp=X(1,:); // vectore x de prueba
        D=X(2:a,:); //matriz x sin 1
        Yc=Y(2:a,:); // matriz y sin 1
    end,
    if k>1 & k<a-1 then
        Dp=X(k,:);
        D1=X(1:k-1,:);
        D2=X(k+1:a,:);
        D=[D1; D2];
        clear D1; clear D2;
        Y1=Y(1:k-1,:);
        Y2=Y(k+1:a,:);
        Yc=[Y1; Y2];
        clear Y1; clear Y2;
    end,
    if k==a-1 then
        Dp=X(a,:);
        D=X(1:a-1,:);
        Yc=Y(1:a-1,:);
    end,
    //-----//
    Xi=inv(D'*D);
    Xs=Xi*D'; clear Xi;
    B=Xs*Yc;
    if k==1 then
        Pred=Dp*B;
    end,
    if k>1 then
        Pred=[Pred; Dp*B];
    end,
end,
//-----//
// Termina la ejecucion de MLR

```

Algoritmo para el reordenamiento de la sección otros resultados

```

l1=list('Ubicación de la variable',2,['Consola','Disco']);
Pregun=x_choices('',list(l1));

```

```

if Pregun(1)==2 then
    rep=uigetfile(['*.txt'; '*.sce'; '*.xls*']);
    cod=str2code(rep);
    clear rep;
    [fd,SST,Sheetnames,Sheetpos]=xls_open(""+code2str(cod)+")
    //lee la primera hoja de datos o espectroscopica
    [data,TextInd] = xls_read(fd,Sheetpos(1))
    //cierra el flujo de la hoja de calculo
    mclose(fd)
    // se ha leído la matriz de datos espectroscopicos
    clear cod;
end,
if Pregun(1)==1 then
    data=data;
end,
//-----//
stacksize(59000000);
l1=list('Espectros a lo largo de',1,['filas','columnas']);
Pregunta=x_choices('',list(l1)); //pregunta como estan ordenados los datos, si los espectros estan
como filas o como columnas
if Pregunta(1)==1 then
    data=data';
end;
[f c]=size(data);
disp('Ya estan leidas las matrices de datos!!');
//-----//
// 4 formas de pre-procesamiento de los datos:
n = x_choose(['1. Covarianza en el origen';'2. Covarianza en el centro';'3. Correlacion en el
origen';'4. Correlacion en el centro';'5. Autoescalado (para LIF)'],['Tipo de Pre-
procesamiento'],'Salir');
//-----//
select n,
case 1 then
    M=data',
    // la matriz de covarianzas en origen
    //-----//
case 2 then
    [f c]=size(data);
    prom=mean(data, 'c');
    // Promedio a cada longitud, o sea, por filas
    // Desviacion estandar a cada longitud, o sea, por filas
    for j=1:c,
        data(:,j)=data(:,j)-prom;
    end;
    clear prom;
    stacksize('max');

```

```

M=data',
// Aqui se ha calculado la matriz de covarianzas centrada
//-----//
case 3 then
[f c]=size(data);
for i=1:f,
    escal(i,1)=norm(data(i,:));
end;
for j=1:c,
    data(:,j)=data(:,j)./escal;
end;
clear escal;
stacksize('max');
M=data',
// calculado la matriz de correlaciones en origen
//-----//
case 4 then
[f c]=size(data);
prom=mean(data, 'c');
// Promedio a cada longitud, o sea, por filas
for i=1:f,
    escal(i,1)=norm(data(i,:)-prom(i));
end;
for j=1:c,
    data(:,j)=(data(:,j)-prom)./escal;
end;
clear escal; clear prom;
stacksize('max');
M=data',
// Matriz de correlaciones centradas
//-----//
case 5 then
[f c]=size(data);
prom=mean(data, 'c');
// Promedio a cada longitud, o sea, por filas
for i=1:f,
    escal(i,1)=sqrt((sum((data(i,:)-prom(i))**2))/(c-1));
end;
for j=1:c,
    data(:,j)=(data(:,j)-prom)./escal;
end;
clear escal; clear prom;
stacksize('max');
M=data',
// Matriz autoescalada
//-----//

```

```

end;
disp('Termina pre-tratamiento!!');
// Termina pre-tratamiento de datos
//-----//
stacksize('max');
[U,S,V] = svd(M);
[f c]=size(S);
//clear V;
score=U*S;// clear U; clear S;
//se han calculado los eigenvectores
//-----//
txt=['Eje x';'Eje y';'Eje z'];
sig=x_mdialog('Factores a tratar por hélice',txt,['2';'3';'1'])
f1=evstr(sig(1))
f2=evstr(sig(2))
f3=evstr(sig(3))
if f1<0 or f1>c then
    if f2<0 or f2>c then
        if f3<0 or f3>c then
            messagebox('Valores escogidos >=0 y <='+string(c)+'');
            txt=['Eje x';'Eje y';'Eje z'];
            sig=x_mdialog('Factores a graficar',txt,['2';'3';'1'])
            f1=evstr(sig(1))
            f2=evstr(sig(2))
            f3=evstr(sig(3))
        end,
    end,
end,
//Termina el programa de PFA (Análisis de Factores Principales)
//Inicia el proceso helicoidal

D=[score(:,f1),score(:,f2)];
[f c]=size(D);
vn=sqrt(sum(D.*D,'c'));
// Hasta aqui se calcula la maxima distancia euclidea del centro coordenado

maxim=max(vn);
normal=maxim*ones(f,1);

for j=1:c,
    temp=D(:,j)./normal;
    D(:,j)=temp;
end,
// Hasta aqui se normalizan las coordenadas x e y

X=D(:,1); Y=D(:,2); Z=score(:,f3);

```

```

for i=1:f,
  if X(i)>0 then
    if Y(i)<0 then
      tetha(i)=(%pi*2)+(atan(Y(i)/X(i)));
    else tetha(i)=(atan(Y(i)/X(i)));
    end,
  end,
  if X(i)<0 then
    if Y(i)<0 then
      tetha(i)=(%pi)+(atan(Y(i)/X(i)));
    else tetha(i)=(%pi)+(atan(Y(i)/X(i)));
    end,
  end,
  if i==f then
    disp(' Trabajo Terminado...');
  end,
end,
plot(Z,tetha,'*g');
xgrid(2);
xlabel('Tetha vs. z');
// Donde finalmente se ha calculado el angulo con su conexión

```

Algoritmo para hacer la derivada

```

//-----DERIVADA POR DIFERENCIAS DIVIDIDAS-----//
//Este programa sirve para calcular la derivada de un vector
//usando diferencias divididas
//Funciona con el filtro de Fourier antes de ejecutar este
//programa
//Se debe tranformar X=real(xx1), y este funcionará perfecto
X=real(xx1);
[f c]=size(X);
x=[1:f]';
//-----//
//comienza la derivacion del espectro
for i=2:f-1,
  if i==2 then
    der(i-1)=0;
  end,
  der(i)=(1/2)*(((X(i+1)-X(i))/(x(i+1)-x(i)))+(X(i)-X(i-1))/(x(i)-x(i-1))));
  if i==(f-1) then
    der(i+1)=0;
  end,
end,
// termina la derivacion
//-----//

```

```

figure()
subplot(2,2,1);
plot(x,X);
xlabel('Funcion original');
subplot(2,2,2);
plot(x,der);
xgrid(2);
xlabel('Derivada de la funcion');
// se han graficado los resultados
//-----//
//comienza la busqueda de los ceros de la derivada del espectro
k2=0;
for i=2:f-2,
    if der(i)<0 & der(i+1)>0 then
        k2=k2+1;
        pv2(k2)=i;
    elseif der(i)>0 & der(i+1)<0 then
        k2=k2+1;
        pv2(k2)=i;
    end,
end,
//-----//
//Ha terminado la busqueda de los ceros de la derivada del espectro
//Comienza derivada de la derivada
for i=3:f-2,
    if i==3 then
        derde(i-1)=0; derde(i-2)=0;
    end, derde(i)=(1/2)*(((der(i+1)-der(i))/(x(i+1)-x(i)))+(der(i)-der(i-1))/(x(i)-x(i-1))));
    if i==(f-2) then
        derde(i+1)=0; derde(i+2)=0;
    end,
end,
// termina la derivacion de la derivada
//-----//
//comienza la busqueda de maximos y minimos de la segunda derivada
k1=0;
for i=4:f-3,
    pen1=derde(i)-derde(i-1); pen2=derde(i+1)-derde(i);
    if pen1<0 & pen2>0 then
        k1=k1+1;
        pv1(k1)=i;
    elseif pen1>0 & pen2<0 then
        k1=k1+1;
        pv1(k1)=i;
    end,
    if i==f-3 then

```

```

    k1=k1+1;
    pv1(k1)=f;
end,
end,
//-----//
//comienza la busqueda de ceros en la segunda derivada
//for i=3:f-3,
//if derde(i)<0 & derde(i+1)>0 then
//k2=k2+1;
//pv2(k2)=i;
//elseif derde(i)>0 & derde(i+1)<0 then
//k2=k2+1;
//pv2(k2)=i;
//end,
//end,
//-----//
// ahora a mostrar las tablas de los resultados
disp('Hay VPs en el espectro!!');
[f1 c]=size(pv1);
[f2 c]=size(pv2);
subplot(2,2,3);
plot(x,derde);
xgrid(2);
xtitle('Derivada de la derivada');
subplot(2,2,4);
plot(x,X);
xgrid(2);
xtitle('VPs');
//for i=1:f1,
// plot(pv1,X(pv1),'*g'); //maximos y minimos en 2-da derivada
//end,
for i=1:f2,
    plot(pv2,X(pv2),'*r'); //ceros en 1-ra y 2-da derivada
end,
//-----//
l1=list('Unificar puntos criticos?',1,['Si','No']);
Pregunta=x_choices(' ',list(l1));
if Pregunta(1)==1 then
    k=0;
    const1=sqrt((f1)*(f**2)); const2=sqrt((f2)*(f**2));
    while (const1~=norm(pv1) & const2~=norm(pv2)),
        [m1,p1]=min(pv1); [m2,p2]=min(pv2);
        if m1<m2 then
            k=k+1;
            pv(k)=m1; pv1(p1)=f;
        elseif m1>m2 then

```

```

    k=k+1;
    pv(k)=m2; pv2(p2)=f;
else m1==m2 then
    k=k+1;
    pv(k)=m1; pv1(p1)=f; pv2(p2)=f;
end,
end,
end,
end,

```

Algoritmo para hacer un pretratamiento

```

//-----//
rep=uigetfile(['*.txt'; '*.sce'; '*.xls*']);
cod=str2code(rep);
clear rep;
[fd,SST,Sheetnames,Sheetpos]=xls_open('+code2str(cod)+')
//lee la primera hoja de datos o espectroscopica
[data,TextInd] = xls_read(fd,Sheetpos(1))
//lee la segunda hoja de datos o de propiedades
[Y,TextInd] = xls_read(fd,Sheetpos(2))
//cierra el flujo de la hoja de calculo
mclose(fd)
// se ha leido la matriz de datos espectroscopicos
clear cod;
//-----//
l1=list('Espectros a lo largo de',2,['filas','columnas']);
Pregunta=x_choices('',list(l1));
if Pregunta(1)==1 then
    data=data';
end;
//para acomodarlos como vectores columna
//-----//
l1=list('Deseas trazar linea base??',1,['Si','No']);
Pregunta3=x_choices('',list(l1));
if Pregunta3(1)==1 then
    [a b]=size(data);
    for i=1:b,
        zurda=data(1:200,i); diest=data(2046:2246,i);
        [a1 b1]=min(zurda); [a2 b2]=min(diest);
        m=(a2-a1)/((b2+2046)-b1);
        // y-y1=m(x-x1)-----y=mx+(y1-mx1)
        inte=a1-m*b1;
        function[y]=lineb(j)
            y=(m*j)+inte
        endfunction
    end
end

```

```

    for j=1:a,
        [y]=lineb(j);
        data(j,i)=(data(j,i))-(y);
    end,
end,
figure()
plot(data);
title('Espectros con linea base lineal');
xgrid;
end,
//Termina la linea base
//-----//
l1=list('Deseas normalizar a longitud unidad??',1,['Si','No']);
Pregunta3=x_choices('',list(l1));
if Pregunta3(1)==1 then
    [a b]=size(data);
    for i=1:b,
        esc(1,i)=norm(data(:,i));
    end,
    for j=1:a,
        data(j,:)=(data(j,:))./esc;
    end,
    figure()
    plot(data);
    title('Espectros Normalizados a longitud unidad');
    xgrid;
end,
//termina normalizacion a longitud unidad
//-----//
l1=list('Deseas normalizar al valor maximo??',1,['Si','No']);
Pregunta3=x_choices('',list(l1));
if Pregunta3(1)==1 then
    [a b]=size(data);
    for i=1:b,
        [mag maxim]=max(data(:,i));
        esc(1,i)=mag;
    end,
    for j=1:a,
        data(j,:)=(data(j,:))./esc;
    end,
    figure()
    plot(data);
    title('Espectros Normalizados al valor maximo');
    xgrid;
end,
//termina normalizacion a longitud unidad

```

```

//-----//
l1=list('Desea Auto-escalamiento??',2,['Si','No']);
Pregunta=x_choices('',list(l1));
if Pregunta(1)==1 then
  [a b]=size(data);
  prom=mean(data, 'c');
  // Promedio a cada longitud, o sea, por filas
  for i=1:a,
    escal(i,1)=sqrt((sum((data(i,:)-prom(i))**2))/(c-1));
  end;
  for j=1:b,
    data(:,j)=(data(:,j)-prom)./escal;
  end;
  figure()
  plot(data);
  title('Espectros Auto-escalados');
end,
//termina auto-escalamiento
//-----//

```