

**DETERMINACIÓN DE LA RELACIÓN HIDRÓGENO / CARBONO DE
COMPUESTOS ORGANICOS MEDIANTE ESPECTROSCOPIA INFRARROJA
CON REFLECTANCIA TOTAL ATENUADA (FTIR-ATR) Y ESPECTROSCOPIA
RAMAN**

CAROLINA QUIROGA PÉREZ

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUIMICA
BUCARAMANGA
2015**

**DETERMINACIÓN DE LA RELACIÓN HIDRÓGENO / CARBONO DE
COMPUESTOS ORGANICOS MEDIANTE ESPECTROSCOPIA INFRARROJA
CON REFLECTANCIA TOTAL ATENUADA (FTIR-ATR) Y ESPECTROSCOPIA
RAMAN**

CAROLINA QUIROGA PÉREZ

Trabajo de grado para optar al título de QUIMICA

**DIRECTOR
ENRIQUE MEJÍA OSPINO
QUÍMICO, Ph.D.**

**Co-DIRECTOR
LESLY VIVIANA MELENDEZ
Química, Msc.**

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

2015

AGRADECIMIENTOS

EN CRISTO QUE ME FORTALECE

Mira que te mando que te esfuerces, y seas valiente; no temas ni desmayes, porque yo el señor tú DIOS soy contigo en donde quiera que fueres.

Josué 1:9

A mi madre por su apoyo incondicional, sus cuidados y su amor y a mi padre que en paz descansa por ser el mejor ejemplo a seguir por haber cumplido a cabalidad la misión dada por Dios de ser papá

A mi hijo por ser mi inspiración, porque es el motor que me impulsa a continuar

Al Doctor Enrique Mejía por su orientación y completa disposición para enseñarme en este proceso de aprendizaje

A todos mis amigos y amigas de la familia LEAM en cabeza del profesor Rafael Cabanzo quienes de una u otra forma me ayudaron a culminar este trabajo y en especial a Lesly Viviana Meléndez Correa por ser ese apoyo incondicional como codirectora de mi tesis y como una gran amiga

Al profesor Jaime Alberto Camargo por ser ese apoyo emocional cuando perdí a mi padre impulsándome a continuar con mi proyecto de grado

Los llevaré siempre en mi corazón... MUCHAS GRACIAS

CONTENIDO

	Pág.
INTRODUCCIÓN	13
1. PLANTEAMIENTO DEL PROBLEMA	15
1.1 JUSTIFICACIÓN	15
2. OBJETIVOS	17
2.1 OBJETIVO GENERAL	17
2.2 OBJETIVOS ESPECÍFICOS	17
3. ESTADO DEL ARTE	18
4. MARCO TEÓRICO	22
4.1 ESPECTROSCOPIA DE INFRARROJO	23
4.1.1 Espectroscopia de infrarrojo con Reflectancia Total Atenuada (IR / ATR)	27
4.1.2 Regiones Espectrales	30
4.1.3 Interpretación de Espectros Infrarrojos	31
4.2 ESPECTROSCOPIA DE RAMAN	35
4.3 QUIMIOMETRÍA	37
4.3.1 Calibración Multivariable.	38
4.3.1.1 Métodos de calibración multivariable	41
4.3.1.2 Construcción de modelos de calibración multivariados	41
4.3.2. Calibración de métodos basados en reducción de variables	43
4.3.2.1. Análisis por componentes principales (PCA)	45
4.3.2.2. Regresión por mínimos cuadrados parciales (regresión PLSR)	46
4.3.3. Validación del modelo	49
4.3.4. Predicción de muestras desconocidas	49
5. METODOLOGÍA	51
5.1 SELECCIÓN DE LAS MUESTRAS ORGÁNICAS	53
5.2 ESPECTROSCOPIA INFRARROJA CON REFLECTANCIA TOTAL ATENUADA (FTIR-ATR)	54
5.2.1 Adquisición de espectros MIR	55
5.2.2 Análisis de espectros MIR de las 13 muestras orgánicas	55

5.3 ESPECTROSCOPIA RAMAN	56
5.3.1 Verificación del desempeño del Equipo	59
5.3.2 Adquisición y Análisis de espectros RAMAN	59
5.3.3 Análisis de los espectros obtenidos	60
5.4 PROCESAMIENTO DE DATOS ESPECTROSCÓPICOS	61
5.4.1 Espectroscopía infrarroja	62
6. ANÁLISIS DE RESULTADOS	66
6.1 ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)	66
6.1.1 Espectroscopia Infrarroja (FTIR-ATR)	66
6.1.2 Espectroscopia Raman	73
6.2 DESARROLLO DEL MODELO DE MÍNIMO CUADRADO PARCIALES (PLS)	78
6.2.1 Calibración del modelo PLS para la espectroscopia Infrarroja (FTIR-ATR)	79
6.2.2 Calibración del modelo PLS para la Espectroscopia Raman	83
7. CONCLUSIONES	89
8. RECOMENDACIONES	91
BIBLIOGRAFÍA	92

LISTA DE TABLAS

	Pág.
Tabla 1. Regiones del espectro infrarrojo	30
Tabla 2. Frecuencias de absorción de infrarrojo de algunas unidades estructurales comunes	32
Tabla 3. Relación H/C en muestras orgánicas usadas para el modelo de calibración	53
Tabla 4. Prueba de repetibilidad del Espectrómetro RAMAN	59
Tabla 5. Desviación estándar para cada normalización con respecto a las longitudes de onda original en los espectros obtenidos por Espectroscopia Infrarroja	63
Tabla 6. Desviación estándar para cada normalización con respecto a las longitudes de onda original en los espectros obtenidos por Espectroscopia Raman	65
Tabla 7. Varianza explicada por los componentes principales de las muestras orgánicas por FTIR-ATR	67
Tabla 8. Número asignado para cada muestra orgánica representados en el gráfico de Scores para espectroscopia infrarroja	68
Tabla 9. Varianza explicada por los componentes principales de las muestras orgánicas por Raman	74
Tabla 10. Varianza explicada por los componentes principales de las muestras orgánicas para PLS	79
Tabla 11. Validación del modelo de predicción	83
Tabla 12. Varianza explicada por los componentes principales de las muestras orgánicas	83
Tabla 13. Validación del modelo de predicción	88

LISTA DE FIGURAS

	Pág.
Figura 1. Relación H/C y obtención de acetileno	19
Figura 2. Diagrama de energía potencial	25
Figura 3. Geometría de una unidad de muestra de reflexión total atenuada (ATR)	28
Figura 4. Aparato de reflectancia total atenuada. (a) Muestra colocada sobre la placa de reflexión; (b) adaptador de reflexión interna.	29
Figura 5. Tipo de vibraciones permitidas para las moléculas	33
Figura 6. Modos normales de vibración para la molécula de Agua	33
Figura 7. Esquema de un Espectroscopio Raman.	37
Figura 8. a) Representación de dos componentes principales, PC_1 y PC_2 , para dos variables, X_1 y X_2 . b) Proyección sobre su eje de los puntos referidos sobre PC_1 y PC_2	46
Figura 9. Descripción gráfica del método de regresión PLS	47
Figura 10. Esquema del desarrollo de la metodología	52
Figura 11. A Espectrómetro FTIR NICOLET iS50 - B. Celda ATR del Espectrómetro	54
Figura 12. Espectros originales infrarrojo	56
Figura 13. Espectrómetro Raman	57
Figura 14. A. Microscopio - B. Porta objetos con laminilla de Sílice	58
Figura 15. A. Montaje para la toma de espectros RAMAN de muestras líquidas - B. Vial con muestra orgánica líquida y tapón de cera de abeja.	58
Figura 16. Espectros Raman Originales de las 12 muestras orgánicas	61
Figura 17. Espectros IR Normalizados	63
Figura 18. Espectros Raman Normalizados	65
Figura 19. Gráfico de Scores de los primeros componentes principales (PC_1 vs PC_2) de las 13 muestras orgánicas por espectroscopia infrarroja	68
Figura 20. Gráfico de Scores de los primeros componentes principales (PC_1 vs PC_2) de las 10 muestras orgánicas sólidas a temperatura ambiente	70

Figura 21. Gráfico de Loadings 2D	71
Figura 22. Gráfico de Loadings lineal	72
Figura 23. Gráfico de Influencia para PC-5 para las 13 muestras orgánicas	72
Figura 24. Gráfico de Leverage para PC5 en el modelo PCA	73
Figura 25. Gráfico de Varianza Explicada en el modelo de PLS para muestras orgánicas	75
Figura 26. Gráfico de Scores de los componentes principales (PC_1 vs PC_2) de las 12 muestras orgánicas	75
Figura 27. Gráfico de Loadings 2D de 12 muestras orgánicas por Espectroscopia Raman	76
Figura 28. Gráfico de Loadings Lineal	77
Figura 29. Gráfico de Influencia para PC-5 de las 13 muestras orgánicas	77
Figura 30. Gráfico de Leverage para PC_5 en el modelo PCA	78
Figura 31. Gráfica de varianza explicada en el modelo para muestras orgánicas	80
Figura 32. Error calculado en función del número de factores principales	81
Figura 33. Curvas de calibración y validación del modelo de predicción de la relación H/C	82
Figura 34. Gráfica de varianza explicada en el modelo para muestras orgánicas	84
Figura 35. Error calculado en función del número de factores principales	85
Figura 36. Coeficientes de regresión	86
Figura 37. Curvas de calibración y validación del modelo de predicción de la relación H/C	87

RESUMEN

TÍTULO

RELACIÓN HIDRÓGENO/ CARBONO DE COMPUESTOS ORGANICOS MEDIANTE ESPECTROSCOPIA INFRARROJA CON REFLECTANCIA TOTAL ATENUADA (FTIR-ATR) Y ESPECTROSCOPIA RAMAN

AUTOR

CAROLINA QUIROGA PÉREZ**

PALABRAS CLAVE

Fórmula empírica, Espectroscopia Raman, Espectroscopía MIR-ATR, relación H/C, compuestos orgánicos.

DESCRIPCIÓN

En el presente trabajo se desarrollaron 2 modelos matemático – estadístico para predecir la relación H/C de compuestos orgánicos mediante Espectroscopía Infrarroja con Reflectancia Total Atenuada (FTIR-ATR) y Espectroscopia Raman. Las 13 muestras analizadas fueron suministradas por los laboratorios de síntesis orgánica y la escuela de química de la Universidad Industrial de Santander y su caracterización se realizó correlacionando las señales espectrales con la relación H/C mediante el de análisis de componentes principales (PCA) que permite la reducción de la dimensionalidad de los datos y la Regresión por mínimos cuadrados parciales (PLS) del cual se obtienen el modelo de predicción o curva de calibración. El beneficio de emplear la técnica Raman está en el estudio de la forma de comportamiento de moléculas de alta simetría, o cromóforos dentro de la molécula, para las cuales no puede obtenerse suficiente información del espectro infrarrojo, siendo ambas técnicas complementarias para estos análisis. Los modelos de predicción se desarrollaron empleando el rango de 4000 a 400 cm^{-1} para Espectroscopía Infrarroja y 100 a 3200 cm^{-1} para Espectroscopía Raman.

El coeficiente de correlación (R^2) entre los valores de calibración y validación por los modelos fue de 0,99 y 0,97 para el modelo de IR y 0,99 y 0,96 para el modelo de Raman. Los errores estándar de calibración (RMSEC) y predicción (RMSEP) obtenidos fueron de 0,02 y 0,012 para el modelo de IR y de 0,04 y 0,012 para el modelo Raman.

Al realizar la validación cruzada de los modelos a través de la determinación del contenido de H/C se encontró que el modelo de mejor desempeño fue realizado mediante el uso de datos de Espectroscopia Infrarroja (FTIR-ATR) según los resultados obtenidos para las dos técnicas espectroscópicas.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Química. Director: Enrique Mejía Ospino. Químico Ph.D.

SUMMARY

TITLE

RELATIONSHIP HYDROGEN / CARBON ORGANIC COMPOUNDS BY INFRARED SPECTROSCOPY WITH ATTENUATED TOTAL REFLECTANCE (ATR-FTIR) AND RAMAN SPECTROSCOPY*

AUTHOR

CAROLINA QUIROGA PÉREZ**

KEYWORDS

EMPIRIC FORMULE, RAMAN SPECTROSCOPY, SPECTROSCOPY MIR-ATR.

DESCRIPTION

In this paper we developed two mathematical models - statistic for predicting the H/C of organic compounds by infrared spectroscopy Attenuated Total Reflectance (ATR-FTIR) and Raman spectroscopy. The 13 samples analyzed were provided by laboratories in organic synthesis and chemistry school of Industrial University of Santander and characterization spectral signals with the H/C was performed correlating by the principal component analysis (PCA) that allows reducing the dimensionality of data and partials Least-Squares Regression (PLS) which the prediction model or calibration curve are obtained. The benefit of the Raman technique is used to study the behavior of molecules form of high symmetry or chromophores within the molecule, for which sufficient information can not be obtained infrared spectrum, both of complementary techniques for these analyzes.

The prediction models were developed using the range of 4000-400 cm^{-1} for Infrared Spectroscopy and 100-3200 cm^{-1} for Raman spectroscopy.

The correlation coefficient (R^2) between the values for calibration and validation was 0.99 models and model 0.97 to 0.99 and 0.96 IR for Raman model. The standard errors of calibration (RMSEC) and prediction (RMSEP) obtained were 0,02 and 0,012 for the IR model and 0,04 and 0,012 for Raman model.

To perform cross-validation of the models through the determination of H / C was found that the model best performance was conducted using data Infrared Spectroscopy (FTIR-ATR) according to the results obtained for the two techniques Spectroscopic.

* Work Degree

** Science Faculty. School of Chemistry. Director: Enrique Mejía Ospino. Químico Ph.D.

INTRODUCCIÓN

Todo organismo viviente está constituido por sustancias orgánicas. Las proteínas que forman el cabello, la piel y los músculos, el ADN que controla la herencia genética, los alimentos que nos nutren, las prendas que nos mantienen calientes y las medicinas que nos curan son sustancias orgánicas. La característica que diferencia las sustancias orgánicas de las demás es que éstas contienen el elemento carbono, que se encuentra en el segundo periodo de la tabla periódica y tiene número atómico 6; es el principal elemento de los compuestos orgánicos y la mayor parte de ellos contiene también hidrógeno; por supuesto que el carbono y el hidrógeno no son los únicos elementos que contienen los compuestos orgánicos. Su espectro es mucho más extenso e incluye:

- Hidrocarburos simples que solamente contienen carbono e hidrógeno como el metano, etano, eteno, benceno, etc.
- Compuestos orgánicos con O, N, S, P y halógenos. Estos compuestos pueden ser clasificados sistemáticamente por sus rasgos estructurales característicos y por su comportamiento químico en mono, di, tri y poli funcionales.
- El último grupo está constituido por innumerables sustancias de origen natural y sintético de uso común y familiares a cualquier individuo, tales como pinturas, fertilizantes, drogas, alimentos, fibras, plásticos, insecticidas, fertilizantes, cosméticos entre otros (1).

De todas las propiedades de un compuesto orgánico como son densidad, conductividad, isometría, reactividad entre otras, la que da más información acerca de su estructura es su espectro infrarrojo y puede emplearse para establecer la identidad de los compuestos y revelar su estructura; así mismo la espectroscopia Raman proporciona información sobre la composición molecular de una muestra que en ocasiones, no se logran evidenciar por Espectroscopia Infrarroja y por lo tanto son técnicas complementarias usadas en este trabajo de investigación. Por sus múltiples propiedades estas técnicas espectroscópicas son fácilmente

aplicables a campos como la geología, biología, química, arqueometría, medicina forense y farmacéutica.

La relación de hidrógeno con respecto a carbono en muestras orgánicas es de vital importancia en el estudio de síntesis orgánica, ya que representa una pequeña parte para la determinación de la fórmula empírica de una sustancia la cual nos revela la estructura molecular de un compuesto. En la Universidad Industrial de Santander se cuenta con métodos gravimétricos y Volumétricos que hacen una determinación cuantitativa en base a la medición de la masa y el volumen respectivamente de compuestos pero las reacciones químicas involucradas en el proceso deben ser rápidas, completas y con estequiometría bien definida para calcular la cantidad de analito presente en una muestra en base a la conservación de la masa y de la carga y así poder establecer la fórmula química de compuestos; todos estos pasos introducen errores en el pretratamiento de las muestras y además errores de tipo instrumental y humano (2).

Actualmente el uso de técnicas espectroscópicas con métodos estadísticos se han estado implementando en la determinación de diferentes propiedades físico químicas de muestras de interés con el fin de mejorar y estandarizar nuevos métodos de caracterización; se realizó este trabajo experimental como una contribución para determinar parte de la fórmula empírica de sustancias orgánicas teniendo el modelo matemático establecido que permita el tratamiento de las señales de los espectros y para ello se emplearon 13 muestras orgánicas usadas en los laboratorios de síntesis orgánica de la Universidad Industrial de Santander las cuales fueron analizadas por Espectroscopia Raman y espectroscopia infrarroja con reflectancia total atenuada. Para el análisis estructural de estas moléculas orgánicas fue necesario realizar dos modelos matemático-estadístico que permitieron determinar la relación Hidrógeno/Carbono de muestras orgánicas en estado líquido y sólido que contienen heteroátomos como oxígeno, nitrógeno entre otros, con características similares a las seleccionadas en el modelo de calibración.

1. PLANTEAMIENTO DEL PROBLEMA

La relación H/C es muy importante en síntesis orgánica y en general en caracterización e identificación de sustancias químicas ya que nos aporta información sobre los elementos que forman el compuesto, la proporción de átomos que tienen y sus propiedades.

Actualmente en la universidad no se encuentra un método rápido y económico con el cual se pueda determinar la composición elemental de compuestos orgánicos y que permita la fácil elucidación de estructuras de moléculas complejas; tampoco equipos altamente especializados para este tipo de análisis que permitan predecir la composición de cualquier especie orgánica con características similares sólo con la obtención de su espectro ni de protocolos a seguir que proporcionen información de manera eficaz y eficiente sobre las características vibracionales fundamentales que se emplean para la elucidación de estructuras moleculares mediante el comportamiento dinámico de moléculas.

Es por este motivo que en este proyecto se plantea el cálculo de la relación H/C a través del uso de La espectroscopia infrarroja con reflectancia total atenuada (FTIR-ATR) y Espectroscopia Raman.

1.1 JUSTIFICACIÓN

Uno de los principales retos de la química es optimizar las técnicas analíticas e instrumentales y los procesos químicos que son utilizados en la industria, ciencia y tecnología de manera que puedan obtenerse productos más elaborados de mejor calidad y a un menor costo y lo más importante que no ocasionen daños considerables al medio ambiente.

La Espectroscopia infrarroja y Espectroscopia Raman son técnicas analíticas que no requiere un pre tratamiento de la muestra con solventes y por lo tanto minimiza costos, reduce errores y daños ambientales, además es un análisis rápido, no destructivo, capaz de determinar todos los elementos de la tabla periódica, tiene pocas limitaciones debido a que se pueden analizar sustancias sólidas, líquidas y gaseosas. Lo que las hace técnicas muy versátiles que no requieren condiciones ambientales especiales y útiles para la detección de componentes en productos importados debido a que no destruye la muestra y puede traspasar envases.

En el área de síntesis orgánica es muy importante el conocimiento de la relación H/C de una sustancia ya que puede ser utilizada para hallar la fórmula empírica para poder caracterizar compuestos, pero el uso de reactivos implica costos e introducción de errores a los análisis, es por esto que es justificable el uso de nuevas técnicas que posibiliten el estudio de las sustancias con el mínimo de errores posibles y a un bajo costo.

Por todas sus ventajas sobre otros métodos y debido a que no se registran análisis elementales de componentes orgánicos empleados en espectroscopia surge la necesidad de estudiar a fondo elementos como hidrógeno y carbono con el propósito de contribuir en la identificación de la fórmula empírica y de esta forma aportar conocimiento en el avance de la ciencia.

Se implementará los métodos de espectroscopia Raman y espectroscopia Infrarroja (FTIR-ATR) y se emplearán métodos quimiométricos para el análisis de los datos que transformen las señales analíticas en información que se pueda utilizar por medio de la estadística y la matemática para predecir parte de la fórmula empírica (relación H/C) de compuestos orgánicos puesto que la determinación de la fórmula empírica es dependiente de la medición de las proporciones relativas de los constituyentes elementales y no de la determinación de las cantidades absolutas, y de esta forma caracterizar familias de compuestos con propiedades similares en el programa Unscramble con el objetivo de minimizar tiempo y dinero.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Realizar un modelo matemático que prediga la relación H/C de compuestos orgánicos por métodos quimiométricos a partir de datos obtenidos por Espectroscopia infrarroja con reflectancia total atenuada (FTIR-ATR) y Espectroscopia Raman.

2.2 OBJETIVOS ESPECÍFICOS

- Adquirir los espectros de muestras orgánicas por Espectroscopia Infrarroja (FTIR-ATR) y Espectroscopia Raman.
- Mediante el uso del software Unscramble, aplicar los pretratamientos adecuados a los datos espectroscópicos como normalización y corrección de línea base.
- Realizar el análisis PCA para reducir la cantidad de datos a partir de combinaciones lineales de las variables originales y conocer las agrupaciones naturales que las muestras presenten.
- Desarrollar un modelo de predicción PLS para el cálculo de la relación H/C

3. ESTADO DEL ARTE

Con el objetivo de analizar el número de investigaciones que se han llevado a cabo sobre la relación H/C de compuestos orgánicos para el cálculo de la fórmula empírica, se realizó una exploración de la producción científica usando la base de datos de la Universidad Industrial de Santander en donde se pudieron encontrar diferentes fuentes publicadas desde 1945 a 2014 referentes a este trabajo investigativo.

Los primeros estudios sobre la estructura de los compuestos fueron realizados por Adolf Windaus en 1903. Empleó el uso de la degradación oxidativa y otras técnicas por las cuales pudo identificar y localizar diversas unidades estructurales (una cadena lateral isodoyl, un doble enlace, un alcohol secundario, cuatro anillos interconectados, etc.) que finalmente lo llevó a proponer su primera fórmula estructural del colesterol, recibiendo en el año de 1928 el Premio Nobel de química por este trabajo, aunque luego resultó que la estructura era incorrecta; y mediante estudios posteriores en el año de 1932 empleando difracción de rayos x se logró establecer correctamente la estructura de la molécula de colesterol (3).

ERWIN BRAND, L. J. Sidel, W. H (1945) hallaron la fórmula empírica de la β -Lactoglobulina en donde se determinaron los aminoácidos presentes en ésta, fraccionando la molécula y separando sus componentes por distintos métodos, analizando el peso molecular de los residuos obtenidos y calculando la composición elemental (en átomos por mol) mediante la suma de los diversos átomos en los componentes individuales; de esta forma se logró concluir en el artículo que la β -Lactoglobulina consiste en 370 aminoácidos dispuestos en 4 sub-unidades o cadenas con 306 enlaces peptídicos en los cuales hay 4 S-S. La fórmula empírica se representó entonces en términos de sus residuos de aminoácidos utilizando símbolos de los primeros aminoácidos para representar la fórmula de la molécula (4).

Nicholas E. Vanderborgh y William T. Ristau hicieron una estimación de la relación Hidrogeno/Carbono de compuestos orgánicos por pirolisis láser en el año 1973. Los datos experimentales se muestran en este artículo para una amplia variedad de hidrocarburos. Los compuestos se enumeraron por la relación de hidrógeno/carbono de la muestra. Se puede observar inmediatamente que dos tipos de compuestos son evidentes, aquellas que dan un rendimiento alto de acetileno y aquellas que dan un porcentaje bajo de acetileno. Se determinó la relación H/C ya que todos aquellos compuestos que tienen dicha relación por debajo de 0,8 obtienen predominantemente acetileno como podemos apreciar en la siguiente figura obtenida del artículo (5).

Figura 1. Relación H/C y obtención de acetileno

Compound	Empirical formula	H/C	A	Methane, %	Ethane ethylene, %	Acetylene, %
Benzo[d,e,f]-phenanthrene	C ₁₆ H ₁₀	0.63	1.00	1.6 ± 0.8	1.2 ± 0.8	97.2 ± 1.5
o-Terphenyl	C ₁₈ H ₁₄	0.78	1.00	0.8 ± 0.5	0.8 ± 0.5	98.4 ± 1.0
m-Terphenyl				1.4 ± 0.5	1.5 ± 0.5	97.2 ± 1.0
p-Terphenyl				0.2 ± 0.1	0.2 ± 0.1	99.6 ± 0.2
Naphthalene	C ₁₀ H ₈	0.80	1.00	1.3 ± 0.5	1.3 ± 0.5	97.4 ± 1.0
Biphenyl	C ₁₂ H ₁₀	0.83	1.00	1.0 ± 0.5	1.0 ± 0.5	98.1 ± 1.0
1,2-Diphenylethane	C ₁₄ H ₁₄	1.00	0.71	4.6 ± 1.8	2.5 ± 0.8	92.9 ± 2.5
Polystyrene	(C ₈ H ₈) _n	1.00	0.63	5.7 ± 1.5	9.0 ± 1.5	85.3 ± 3.0
Durene	C ₁₀ H ₁₄	1.40	0.14	12.8 ± 2.0	8.4 ± 1.8	78.7 ± 2.5
Paraffin	(CH ₂) _n	2.00	0.00	17.2 ± 1.5	39.1 ± 2.0	43.8 ± 3.0

Fuente: Nicholas E. Vanderborgh and William T. Estimation of the Carbon / Hydrogen Ratio of Organic Compounds by Laser Pyrolysis

David L. Windsor and M. Bonner Denton en julio de 1979 realizaron un estudio sobre la determinación de la fórmula empírica con acople inductivo de Plasma por Cromatografía de detección de gases; en el cual se reportan los estudios iniciales que evalúan la capacidad del ICP-AES para determinar fórmulas empíricas de especies de elución de un cromatógrafo de gases teniendo en cuenta que la determinación de la fórmula empírica depende de la medición de las proporciones

relativas de los constituyentes elementales y no de las cantidades absolutas, pudiéndose encontrar muy útil debido que los factores que limitan la técnica como son la inyección cuantitativa y forma de los picos no son de gran importancia. Las fórmulas empíricas se determinaron a partir de composiciones elementales.

Nitin R. Hardas, Peter C. Uden (1999) determinaron fórmulas de estudios empíricos de compuestos clorofluorocarbonados utilizando cromatografía de gases acoplada a la detección de emisión atómica; obteniendo valores en el intervalo de nano-mol y encontrando que las relaciones Cl/C y H/C están más cerca de los valores teóricos que los ratios de F/C. El análisis de mezclas de CFC dio fórmulas empíricas válidas (6).

Varios estudios se han realizado en contorno a predicción de alguna propiedad mediante tratamiento de análisis multivariable entre los cuales tenemos a Daniel T. Berhe, Anders J. Lawaetz, Soren B. Engelsen, Marchen S. Hviid, René Lametsch (2014) investigaron si la espectroscopia Raman en combinación con el análisis de datos multivariable puede ser utilizado para determinar la temperatura del punto final (EPT); factor muy importante para el control de enfermedades transmitidas por alimentos mediante la inactivación de los microorganismos tales como *Listeria Monocytogenes*, *Salmonella sp* entre otras. Se determinaron las mínimas condiciones de procesamiento térmico de cocinado de los productos y se usó la espectroscopia de fluorescencia para apoyar los resultados de la espectroscopia Raman por la medición de los cambios en el entorno de los aminoácidos aromáticos en la carne cocinada.

Para el tratamiento de los espectros Raman y de fluorescencia se empleó análisis de componentes principales (PCA) con el fin de recuperar información sobre las variaciones espectrales sistemáticas en relación a los tratamientos y se empleó regresión PLS para evaluar la relación entre los espectros Raman y la EPT dando

resultados positivos para emplear estas técnicas en el procesamiento de alimentos y control microbiológico (7).

Otros estudios quimiométricos de las técnicas en infrarrojo y espectroscopia Raman se han utilizado como una herramienta útil para el análisis forense de pinturas como el trabajo realizado por Cyril Muehlethaler, Geneviene Massonnet, Pierre Esseiva (2010) quienes evaluaron las capacidades y limitaciones de los métodos quimiométricos (PCA, HCA) y otros tratamientos matemáticos aplicados en los datos espectroscópicos para estudiar y distinguir 34 muestras de pinturas observando seis grupos correspondientes a las diferentes composiciones de pigmentos logrando disminuir los errores causados en cuanto a la clasificación de estas pinturas y ahorro de tiempo en dichos análisis (8).

4. MARCO TEÓRICO

La “estructura” es la clave de todo en química. Las propiedades de una sustancia dependen de los átomos que contiene y de la forma que estos están conectados. Cuando se conoce la fórmula estructural de un compuesto se pueden predecir además algunas de sus propiedades como son su polaridad, su estado físico, el cálculo de la proporción en peso de los distintos elementos que lo conforman entre otros (9).

En las moléculas orgánicas se pueden diferenciar dos partes esenciales: la cadena de átomos de carbono y los átomos de otros elementos que se unen a ella. Estos átomos son en su mayoría átomos de hidrógeno, asociación que da origen a un grupo numeroso de sustancias orgánicas llamadas hidrocarburos, que se constituyen en la columna vertebral sobre la cual se ensamblan todo los demás compuestos orgánicos. Sin embargo hay otros compuestos como el oxígeno y el nitrógeno que se unen a este esqueleto carbonado mediante diferentes mecanismos químicos, aumentando aún más la gran diversidad de compuestos orgánicos que hoy se conocen.

Es de admirar que el mismo tipo de átomos o agrupación de átomos puede ocupar distintas posiciones dentro de una misma cadena carbonada, lo cual es la razón suficiente para modificar el comportamiento químico de cada una de las moléculas resultantes.

La mayor parte de los compuestos orgánicos requieren un enfoque más sistemático para escribir sus fórmulas estructurales. Se comienza como primer paso con la fórmula molecular que indica que átomos y cuántos de cada uno están presentes en el compuesto. A partir de la fórmula molecular se calcula el número de electrones de Valencia siendo éste el segundo paso; y como último paso se

debe establecer una estructura parcial que muestre el orden en que los átomos están conectados y casi siempre se determina de manera experimental (9).

Las formulas empíricas representan de forma abreviada las sustancias químicas, son determinadas a partir de datos experimentales sintetizando una muestra pura y analizando el porcentaje en peso de cada elemento presente que contiene ese compuesto; (1) no necesariamente coincide con la fórmula molecular. Si no coinciden, la fórmula molecular es un múltiplo de la fórmula empírica

En química orgánica es usual encontrar dos o más compuestos diferentes con la misma fórmula molecular pero con conectividades diferentes como un ejemplo de ellos podemos citar al etanol con el éter dimetílico los cuales son compuestos diferentes, que tienen diferentes propiedades pero la misma fórmula molecular (C_2H_6O) también conocidos como isómeros.

4.1 ESPECTROSCOPIA DE INFRARROJO

Los espectros de absorción, emisión y reflexión en el infrarrojo, de especies moleculares, se pueden explicar asumiendo que todos son el resultado de los distintos cambios energéticos producidos en las transiciones de las moléculas de unos estados de energía vibracionales y rotacionales a otros (10). La molécula, al absorber la radiación infrarroja, cambia su estado de energía vibracional y rotacional. Las transiciones entre dos estados rotacionales requieren muy poca energía, por lo que solo es posible observarlas específicamente en el caso de muestras gaseosas. En el caso del estudio del espectro infrarrojo (IR) de muestras sólidas y líquidas sólo se tienen en cuenta los cambios entre estados de energía vibracional (11).

Podemos explicar los espectros vibracionales mediante la expresión algebraica (1) según la ley de Hook, la fuerza restauradora F es proporcional al desplazamiento

considerando el caso de tener una única masa unida a un resorte que cuelga de un objeto inmóvil.

$$\text{Fuerza} = -k\Delta x \quad (1)$$

Donde k es una constante y Δx la distancia recorrida desde la posición de equilibrio, x_0 . El signo negativo indica que la fuerza tiene dirección opuesta a la posición del enlace químico respecto a la posición de equilibrio: cuando se comprime el enlace, la fuerza que aparece tiende a estirarlo, cuando se estira, aparece una fuerza de compresión.

Un grupo de masas unidas mediante enlaces tiende a asentarse de manera natural con un movimiento específico cuando se agitan. Las masas vibran con una frecuencia que depende de las masas y de la fortaleza del muelle. Estas vibraciones, que aparecen con sus Frecuencias características, son las que medimos en los espectros vibracionales (12).

Para las moléculas diatómicas la vibración se describe mediante la ecuación (2).

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (2)$$

Donde:

ν es la frecuencia natural de vibración de las masas.

k es la constante de fuerza del enlace químico e indica la rapidez de cambio de la fuerza cuando ocurre una compresión o elongación (así un valor alto de k indica enlace fuerte).

μ es la masa reducida que se representa mediante la ecuación (3).

$$\frac{1}{\mu} = \frac{1}{M_1} + \frac{1}{M_2} \quad \text{o} \quad \mu = \frac{M_1 M_2}{M_1 + M_2} \quad (3)$$

La ecuación final al sustituir la ecuación número 3 en la ecuación 2 es la siguiente.

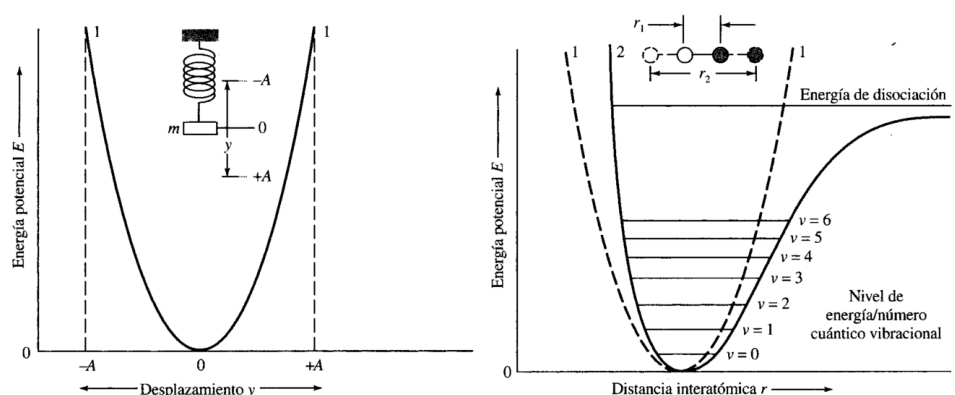
$$v = \frac{1}{2\pi c} \sqrt{\frac{k(m_1+m_2)}{m_1 \times m_2}} \quad (4)$$

Matemáticamente, la energía de los estados vibracionales en una molécula diatómica se puede describir por el modelo del oscilador armónico (figura 1).

Cuando una molécula está constituida por más de dos átomos, las frecuencias vibracionales no se pueden describir con una ecuación tan sencilla como ecuación (2). Las vibraciones provienen de movimientos simultáneos de más de dos átomos y se denominan modos normales.

A partir de las ecuaciones anteriores también podemos deducir dos cosas: la primera, que los enlaces fuertes al tener una mayor constante de fuerza vibran a mayor frecuencia (menor longitud de onda) que los débiles y la segunda es que los enlaces entre átomos con masas grandes vibran a menores frecuencias (mayores longitudes de onda) que los enlaces entre átomos ligeros.

Figura 2. Diagrama de energía potencial



Fuente: Skoog, D. A., Holler, J. H., Nieman, T. A. Principios de Análisis Instrumental

Las transiciones en las que $\Delta v = \pm 1$, originan líneas o bandas de absorción fundamental. La frecuencia de radiación que producirá este cambio es idéntica a la frecuencia de vibración fundamental del enlace (13).

El número de onda del pico de absorción entonces se puede calcular haciendo la aproximación de que el comportamiento de una vibración molecular es análogo al de un oscilador armónico simple, debido a la transición del estado vibracional fundamental al primer estado excitado para diferentes tipos de enlace.

A través de la combinación de las ecuaciones 5 y 6 que representan la variación de la energía potencial como la fuerza multiplicada por la distancia; se obtiene la ecuación 7 que al ser integrada entre la posición de equilibrio ($y=0$) e y forma la ecuación 8 que representa una curva de energía potencial para una oscilación armónica simple. Es una parábola como la representada en la **figura 1** en la cual se puede observar que cuando el resorte está extendido o comprimido a su máxima amplitud A , la energía potencial alcanza un valor máximo y cuando está en equilibrio disminuye a cero

$$dE = -Fdy \quad (5)$$

$$dE = kydy \quad (6)$$

$$\int_0^E dE = k \int_0^y ydy \quad (7)$$

$$E = \frac{1}{2}ky^2 \quad (8)$$

La técnica de Espectroscopia de infrarrojo permite la identificación de grupos funcionales de un espectro mediante la absorción de la radiación por la molécula produciéndose una vibración intramolecular con frecuencia igual a la de la radiación y esto genera señales con frecuencias que corresponden a la vibración de un enlace específico. A continuación se explicará el fenómeno de reflectancia total atenuada en espectroscopia de infrarrojo.

4.1.1 Espectroscopia de infrarrojo con Reflectancia Total Atenuada (IR / ATR)

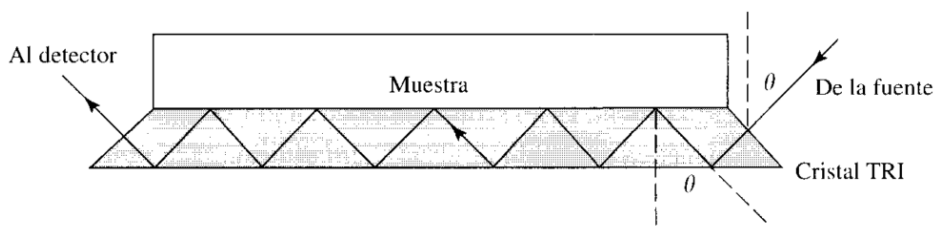
La espectroscopia de reflexión interna es una técnica que permite la obtención de espectros de infrarrojo de muestras que presentan alguna dificultad, como sólidos de limitada solubilidad, películas, fibras, pastas, adhesivos y polvos.

Un accesorio de ATR opera por la medida de los cambios que ocurren en un haz infrarrojo reflejado internamente cuando interactúa con cualquier material absorbente. Un haz IR es direccionado al interior de un cristal ópticamente denso de alto índice de refracción, a un ángulo dado. La reflectancia interna crea una onda evanescente que se proyecta ortogonalmente al interior de la muestra que está en contacto íntimo con el cristal ATR (14).

Teórica y experimentalmente se ha demostrado que durante el proceso de reflexión el haz se comporta como si penetrase una cierta distancia en el medio menos denso antes de reflejarse. La profundidad de penetración, que puede variar desde una fracción de longitud de onda a varias longitudes de onda, depende de la longitud de onda de la radiación incidente, del índice de refracción de los dos materiales, y del ángulo que forma el haz incidente con la interface. La radiación que penetra se denomina onda evanescente (10).

La onda evanescente que alcanza el exterior de la superficie donde se produce la reflexión total interna, interacciona con el material en el lado exterior de la superficie. Si el lado exterior de la superficie absorbe luz a la longitud de onda que se está propagando, la luz que viaja internamente a través del material está atenuada, y se puede registrar el espectro del material inferior. Este tipo de muestreo es muy efectivo para sólidos y líquidos, especialmente en las regiones del infrarrojo medio y del infrarrojo cercano. El instrumento utilizado para esta espectrometría de reflexión total atenuada (ATR) se representa en la siguiente figura (12).

Figura 3. Geometría de una unidad de muestra de reflexión total atenuada (ATR)



Fuente: Rubinson, K. A., Rubinson, J. F. Análisis Instrumental.

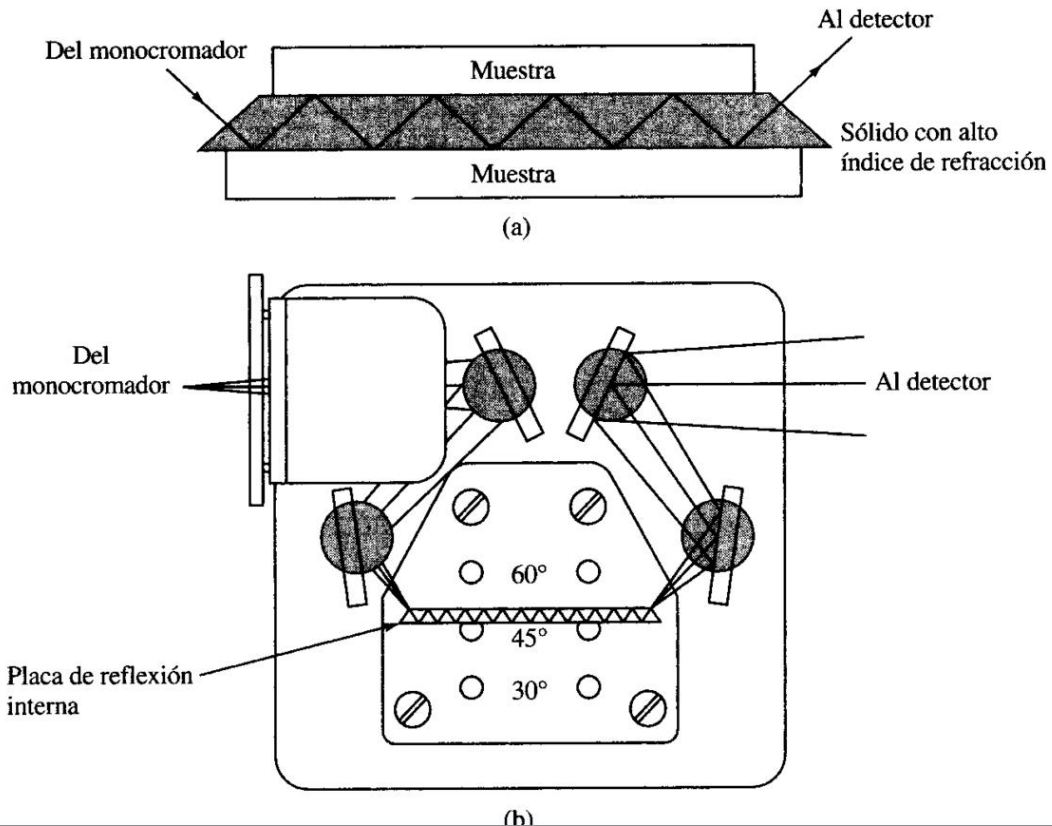
La luz incidente choca con la superficie de la muestra al ángulo elegido que es el mismo que el ángulo que corta el final del cristal. La luz se sitúa para que el rayo incidente choque perpendicularmente al final del cristal (una geometría que minimiza la luz reflejada o cambia sus propiedades de polarización). La luz se propaga por reflexión total interna (TRI). La muestra podría colocarse también en la superficie inferior. Los cristales cilíndricos y las fibras ópticas pueden tener muestras en sus alrededores (12).

Si el medio menos denso absorbe la radiación evanescente, se produce una atenuación del haz en las longitudes de onda de las bandas de absorción, este fenómeno se conoce como Reflectancia Total Atenuada (ATR).

La espectroscopía de reflectancia total atenuada es una de las técnicas espectroscópicas más versátiles y de mayor aplicación ya que se ha aplicado a muchas sustancias como polímeros, gomas y otros sólidos, con ésta técnica se han caracterizado e identificado materiales, se han analizado productos farmacéuticos y de síntesis, se han analizado contaminantes y es usado también en la ciencia forense, biomedicina, agricultura y alimentación entre otros. Los espectros resultantes están libres de los frentes de interferencia, en general se observan los mismos picos con respecto a los espectros de absorción ordinarios, pero sus intensidades relativas son distintas. Una de las mayores ventajas de la espectroscopia de reflectancia total atenuada es que, con una mínima

preparación, se puede obtener fácilmente los espectros de absorción de una gran variedad de tipos de muestras (10) a continuación se observa en la **figura 4** el dispositivo de reflectancia total atenuada y se evidencia claramente el fenómeno.

Figura 4. Aparato de reflectancia total atenuada. (a) Muestra colocada sobre la placa de reflexión; (b) adaptador de reflexión interna.



Fuente: Skoog, D. A., Holler, J. H., Nieman, T. A. *Principios de Análisis Instrumental*.

La información presente en un espectro MIR puede ser usada para estimar la concentración de un componente o para estimar una propiedad física cuando estas reflejan cambios significativos en las características espectrales generadas por la muestra. Para lograr esto, es necesario recurrir a diversos métodos multivariados de análisis que se encuentran agrupados en una rama de la química analítica. La quimiometría es una técnica matemática estadística usada para

extraer información relevante de datos analíticos, en este caso de la información espectral obtenida en la región del infrarrojo medio (15).

4.1.2 Regiones Espectrales. La región del espectro infrarrojo abarca la radiación con números de onda comprendidos entre 12.800 y 10 cm^{-1} , que corresponden a longitudes de onda de 0,78 a 1000 μm^1 . Tanto desde el punto de vista de las aplicaciones como de la instrumentación, es conveniente dividir el espectro infrarrojo en tres regiones denominadas cercano, medio y lejano (10).

En un principio, la región espectral del infrarrojo medio era usada en su mayor parte para el análisis orgánico cualitativo y la determinación estructural, teniendo como base los espectros de absorción. Ahora, como contraste, la espectroscopia en el infrarrojo medio se está comenzando a utilizar además en el análisis cuantitativo de muestras complejas, mediante espectroscopia de absorción y emisión. También han empezado a aparecer aplicaciones de esta región espectral en los estudios microscópicos de superficies, análisis de sólidos mediante reflectancia total atenuada y reflectancia difusa, medidas foto acústica y otras (16). La **tabla 1** muestra el rango en el espectro electromagnético al que corresponde cada región del espectro infrarrojo.

Tabla 1. Regiones del espectro infrarrojo

Región	Intervalo de longitud de onda (λ), μm	Intervalo de número de onda ($\bar{\nu}$), cm^{-1}	Intervalo de frecuencias (ν), Hz
Cercano	0,78 a 2,5	12.800 a 4.000	$3,8 \times 10^{14}$ a $1,2 \times 10^{14}$
Medio	2,5 a 50	4.000 a 200	$1,2 \times 10^{14}$ a $6,0 \times 10^{12}$
Lejano	50 a 1.000	200 a 10	$6,0 \times 10^{12}$ a $3,0 \times 10^{11}$
La más utilizada	2,5 a 15	4.000 a 670	$1,2 \times 10^{14}$ a $2,0 \times 10^{13}$

Fuente: SKOOG,A.Douglas. Principios de análisis Instrumental.




4.1.3 Interpretación de Espectros Infrarrojos. La región Infrarroja (IR) es la porción del espectro electromagnético entre las microondas y la luz visible y cubre el intervalo que va de un poco arriba del visible (7.8×10^{-7} m) a unos 10^{-4} m, pero en química orgánica sólo se usa la región media, desde 2.5×10^{-6} hasta 2.5×10^{-5} m. Por lo general, las longitudes de onda en la región infrarroja se expresan en micrómetros o micras ($1\mu\text{m} = 10^{-6}$ m), y las frecuencias, en números de onda, no en Hertz. El número de onda se da en cm^{-1} y es el recíproco de la longitud de onda, en centímetros:

La radiación electromagnética en la región de 4000 a 625 cm^{-1} corresponde a la separación entre estados de energía vibratoria adyacentes en moléculas orgánicas.

Entre los compuestos orgánicos escogidos en este trabajo tenemos el hexano que aparece como una serie de picos de absorción de forma intensiva y variables. Casi todos los compuestos orgánicos exhiben un pico o grupo de picos cerca de 3000 cm^{-1} debido al alargamiento carbono- hidrógeno. Los picos a 1460 , 1380 y 725 cm^{-1} se deben a diversas vibraciones de torsión.

Al usar espectroscopia de IR para la determinación de estructuras, por lo general se destacan picos en el intervalo de 4000 a 1600 cm^{-1} porque ésta es la región en la cual se encuentran las vibraciones características de los grupos funcionales particulares. La región de 1300 a 625 cm^{-1} se conoce como la región de huellas digitales; es aquí donde el patrón de picos varía más de un compuesto a otro. En la **tabla 2** se listan las frecuencias (en números de onda) asociadas con una variedad de grupos encontrados por lo común en compuestos orgánicos (9).

Tabla 2. Frecuencias de absorción de infrarrojo de algunas unidades estructurales comunes

Stretching vibrations			
<i>Single bonds</i>		<i>Double bonds</i>	
—O—H (alcohols)	3200–3600		1620–1680
—O—H (carboxylic acids)	2500–3600		
 N—H	3350–3500	Aldehydes and ketones	1710–1750
<i>sp</i> C—H	3310–3320	Carboxylic acids	1700–1725
<i>sp</i> ² C—H	3000–3100	Acid anhydrides	1800–1850 and 1740–1790
<i>sp</i> ³ C—H	2850–2950	Acyl halides	1770–1815
		Esters	1730–1750
<i>sp</i> ² C—O	1200	Amides	1680–1700
<i>sp</i> ³ C—O	1025–1200		
		<i>Triple bonds</i>	
		—C≡C—	2100–2200
		—C≡N	2240–2280
Bending vibrations of diagnostic value			
<i>Alkenes</i>		<i>Substituted derivatives of benzene</i>	
Cis-disubstituted	665–730	Monosubstituted	730–770 and 690–710
Trans-disubstituted	960–980	Ortho-disubstituted	735–770
Trisubstituted	790–840	Meta-disubstituted	750–810 and 680–730
		Para-disubstituted	790–840

Fuente: ROBERT C. Atkins.Francis A. Carey. Organic Chemistry., Student Solutions Manual 5ª edición.

Todas las moléculas tienen una cierta cantidad de energía distribuida en su estructura, lo que hace que los enlaces se estiren y contraigan, que los átomos oscilen y que se presenten otras vibraciones moleculares es por esta razón que una molécula orgánica absorba algunas longitudes de onda de la radiación infrarroja pero otras no. En la **figura 5** se representan algunos tipos de vibraciones permitidas.

Figura 5. Tipo de vibraciones permitidas para las moléculas

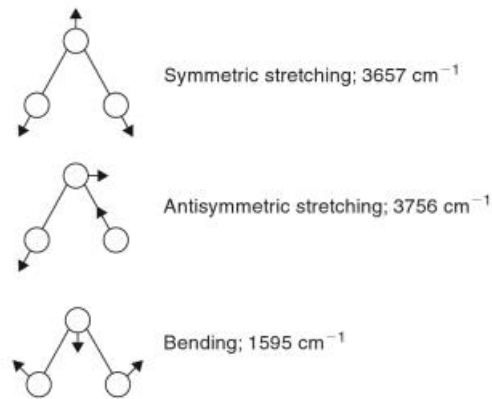
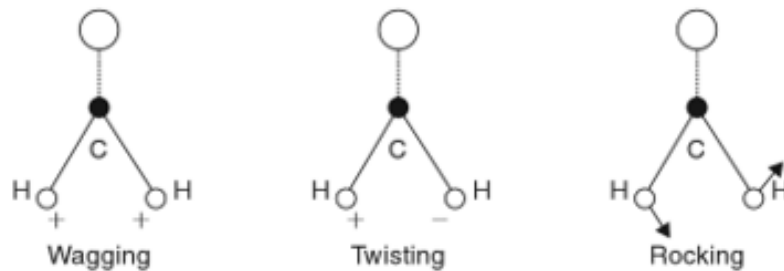


Figura 6. Modos normales de vibración para la molécula de Agua



Fuente: Infrared Spectroscopy for food quality analysis and control. DA-WEN Sun. USA .2009.

Algunas frecuencias vibratorias tienen prohibido aparecer en el espectro infrarrojo por la acción de las reglas de selección. Estas reglas de selección son más restrictivas para las moléculas altamente simétricas. El requisito general para la actividad de infrarrojos de una vibración es que la vibración debe producir un cambio periódico en el momento dipolar; si no se produce ese cambio, la vibración está prohibida en el infrarrojo. Por supuesto, la molécula aún puede llevar a cabo la vibración, pero no se activará por la absorción de radiación infrarroja, y por lo tanto no será detectado en el espectrofotómetro de infrarrojos. (17)

Enlaces C-H:

Los enlaces CH, que se encuentran en grandes cantidades en moléculas orgánicas, muestran estiramiento vibraciones entre 2750 y 3320 cm^{-1} en la región MIR. La ubicación de estas bandas está relacionado con la hibridación de carbono. Como moléculas alifáticas saturadas se caracterizan por bandas de absorción a aproximadamente 3000 cm^{-1} , grupos vinílicos y acetilénicos presentes bandas de absorción a 3100 cm^{-1} y 3300 cm^{-1} , respectivamente.

La presencia de metilo o función de metileno puede evaluarse por la observación entre 1465 y 1370 cm^{-1} de las vibraciones de flexión de C-H de bonos se caracteriza por una banda de absorción relativamente intensa entre 650 y 1000 cm^{-1} . La vibración de tensión del doble enlace C=C (no conjugado) se ha observado entre 1640 y 1666 cm^{-1} .

En la región NIR, el primero y segundo armónicos para las vibraciones de estiramiento de C-H se observan en alrededor de 1700 y 1200 cm^{-1} respectivamente. Bandas de combinación que involucran estiramiento y flexión de C-H ser identificados entre 2000 y 2500 nm y, con menor intensidad, entre 1300 y 1440 nm. (18)

Una de las moléculas de este trabajo investigativo es el hexano el cual presenta un estiramiento vibracional que se observa entre 2850 y 2962 cm^{-1} son característicos de carbonos sp^3 y permiten la identificación de los grupos metilo y metileno. Para grupos metilo, la vibración asimétrica CH_3 se encuentra en 2962 cm^{-1} para hexano y en 2956 cm^{-1} , respectivamente. El grupo metileno hexano muestra una CH_2 asimétrico vibración de estiramiento en 2926 cm^{-1} , así como una CH_2 vibración simétrica a 2864 cm^{-1} .

4.2 ESPECTROSCOPIA DE RAMAN

C.V.Raman, un físico hindú, descubrió que la longitud de onda de una pequeña fracción de la radiación dispersada por ciertas moléculas, difiere de la del haz incidente y que los desplazamientos de la longitud de onda dependen de la estructura química de las moléculas responsables de la dispersión. (12)

Los estudios de Raman se basaron en los realizados con anterioridad por Rayleigh el cual afirmaba que el color azul del mar no es más que el azul del cielo visto en reflexión. Raman realizó un experimento sencillo mediante el que pudo demostrar que el color azul del agua procedía de un fenómeno propio, posteriormente explicado como la dispersión de la luz debido a su interacción con moléculas del agua.

La espectroscopia Raman es una técnica fotónica de alta resolución que proporciona en pocos segundos información química y estructural de casi cualquier material o compuesto orgánico y/o inorgánico permitiendo así su identificación. El análisis mediante espectroscopia Raman se basa en el examen de la luz dispersada por un material al incidir sobre él un haz de luz monocromático lo cual implica una colisión inelástica de un fotón con una molécula, con la transferencia de una parte de la energía del fotón a la molécula, resulta una excitación vibracional; (19) Una pequeña porción de la luz es dispersada inelásticamente experimentando ligeros cambios de frecuencia que son característicos del material analizado e independientes de la frecuencia de la luz incidente. Se trata de una técnica de análisis que se realiza directamente sobre el material a analizar sin necesitar éste ningún tipo de preparación especial y que no conlleva ninguna alteración de la superficie sobre la que se realiza el análisis, es decir, es no-destructiva. (20)

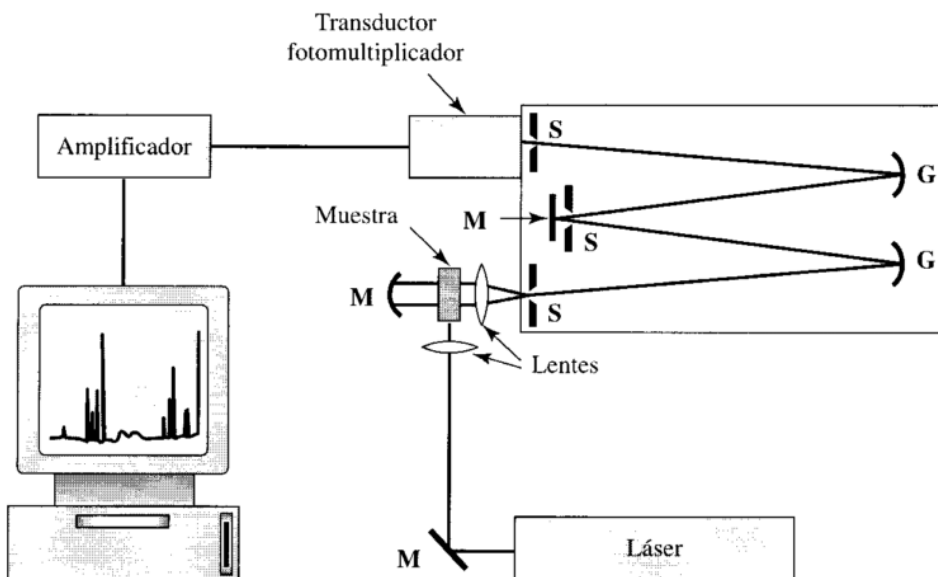
La teoría de la dispersión Raman demuestra que el fenómeno con el mismo tipo de cambios vibracionales cuantizados que se producen en la absorción infrarroja. Por tanto, la diferencia de longitud de onda entre la radiación visible incidente y la dispersada corresponde a las longitudes de onda de la región del infrarrojo medio. En efecto, para una especie determinada, el espectro de dispersión Raman y el espectro de absorción en el infrarrojo, con frecuencia, suelen parecerse mucho. Sin embargo, hay suficientes diferencias entre los tipos de grupos funcionales que son activos en el infrarrojo y los que lo son en Raman, como para que ambas técnicas no resulten competitivas sino complementarias. Para resolver algunos problemas, el infrarrojo es una herramienta superior, sin embargo para otros el Raman proporciona unos espectros más útiles. (10)

La espectroscopia Raman, donde la luz en la región visible se utiliza para excitar a las moléculas de la muestra, tiene requerimientos similares a los necesarios en la absorción óptica y la espectroscopia por fluorescencia. La geometría de los instrumentos Raman es similar a la de un espectroscopio de luminiscencia. La mayor diferencia es que la fuente de la luz, generalmente un láser, es intensa y se puede dirigir o enfocar hacia un pequeño punto de la muestra.

Los instrumentos Raman más modernos utilizan ópticas de volumen holográfico para prevenir la dispersión Rayleigh, y tienen un amplio uso debido a su simplicidad cuando la resolución es tan alta que no se necesita un monocromador doble. Los instrumentos de transformadas de Fourier para Raman se han desarrollado para usarse como excitación en la zona cercana al infrarrojo; de forma parecida ocurre con los instrumentos de infrarrojos. La luz cercana al infrarrojo no excita la fluorescencia interferente ni fotoliza las muestras más inestables. En **la figura 7** se puede apreciar el esquema de un espectroscopio Raman con sus componentes principales usando láseres como fuentes debido a que presentan una alta intensidad para producir la dispersión Raman logrando medir una señal sin confundirla con el ruido del equipo. Además se disponen

espejos que son utilizados como colimadores y un monocromador que proporciona una mayor resolución y el rechazo de longitudes de onda no deseadas

Figura 7. Esquema de un Espectroscopio Raman.



Fuente: Rubinson, K. A., Rubinson, J. F. *Análisis Instrumental*.

4.3 QUIMIOMETRÍA

La quimiometría utiliza métodos matemáticos multivariados para correlacionarlos con alguna propiedad medida de un grupo de muestras, extrae la información más útil y relevante a partir de datos producidos en experimentos Fisicoquímicos (21). El desarrollo de la quimiometría como herramienta de análisis, clasificación y calibración multivariable, se da por la imposibilidad de describir y modelar sistemas químicos mediante la estadística univariada, es una herramienta útil y versátil debido a que transforma una gran cantidad de datos experimentales en información de utilidad para los usuarios. Sus primeras aplicaciones se dieron a finales de la década de 1970 en el área de la química analítica con el fin de analizar datos dependientes de más de una variable simultánea.

En la mayoría de los casos se requiere utilizar todas o gran parte de las variables espectrales puesto que se necesita suficiente información para poder realizar el análisis multivariado que no es más si no el desarrollo de un procedimiento analítico dado (22).

4.3.1 Calibración Multivariable. La selección del método de calibración una vez que se ha aplicado el pre-tratamiento de elección, el modelo de calibración se estructura sobre la base de sus relaciones entre la señal analítica y la propiedad que va a ser determinada (Relación H/C) o de una relación empírica. Una variedad de algoritmos matemáticos están disponibles para la construcción de modelos y existe una amplia gama de técnicas estadísticas para su evaluación y optimización.

Los espectrofotómetros actuales permiten espectros que contienen miles de variables para cada muestra que se obtiene haciendo extenso el trabajo; los métodos de calibración de reducción de variables (PCR; PLSR) son las mejores opciones, no sólo porque reducen las variables, sino también porque suprimen el ruido.

La elección de la gama espectral más adecuado para el desarrollo de un modelo de calibración no es una tarea fácil y con frecuencia se trata de una secuencia sin fin de ejecuciones de prueba y error hasta que se consigue una capacidad predictiva adecuada. Cuando los espectros para el analito objetivo y sus posibles interferentes en forma pura están disponibles, uno puede elegir el rango donde el analito exhibe bandas sustanciales y no incluyen aquellos en los que los interferentes se absorban. Sin embargo, este enfoque es inútil en la región NIR, donde las bandas son típicamente estrechas y fuertemente solapados, y la señal del analito está por lo tanto fácilmente oculta por la señal de la matriz de la muestra. En esta situación, uno puede simplemente calcular la correlación entre la

absorbancia a cada longitud de onda y la cantidad de destino con el fin de trazar el vector resultante en contra de la variable independiente. Esos intervalos que exhiben la correlación más fuerte pueden ser de ayuda con el fin de seleccionar una gama apropiada para desarrollar el modelo. Este enfoque es utilizado por algunos paquetes de software para construir modelos de espectros continuos (23). Una amplia variedad de algoritmos matemáticos para el desarrollo de modelos de calibración están disponibles, con diferentes requisitos de aplicación, pero también con diferente capacidad para modelar las variables independientes, para la obtención de modelos con una buena capacidad predictiva de las muestras. Elegir el número óptimo de factores, componentes principales (PCR) o variables latentes (PLSR) para la definición de un modelo es un paso clave en cualquier proceso de calibración que implica la reducción de variables. Un número de este tipo puede ser seleccionado de varias maneras, la mayoría de los cuales se basan en el error de predicción obtenido con números variables. Con un pequeño número de muestras, el método de elección es por lo general de validación cruzada, que utiliza muestras del conjunto de calibración para comprobar la bondad de ajuste del modelo. Por lo tanto, el conjunto de calibración se divide en varios bloques o segmentos y el modelo se construye utilizando un bloque de datos como segmentos a fin de verificar los resultados y los otros segmentos para desarrollar el modelo; de esta manera, un segmento se deja fuera cada vez. El proceso se realiza para cada factor individual y la salida se acumula con el fin de obtener una estimación fiable de la capacidad predictiva de las muestras de calibración (18).

La calibración multivariable se puede definir como la actividad de encontrar relaciones entre una o más variables de respuesta “y” y una matriz de variables predictorias “x”, de tal forma que se cumpla,

$$y = g(x) \quad (9)$$

La variable “y” puede ser un parámetro cuantitativo o cualitativo que representa la propiedad de interés en el sistema, y la matriz “x” contiene información relevante de la muestra y es determinada por un método de análisis químico o

espectroscópico. La función “g(x)” puede ser difícil de determinar por el número de variables en la matriz “x”. La forma de la función “g (x)” depende del método de regresión empleado, por lo cual puede existir más de una posibilidad de ajuste de los datos diferenciándose principalmente en la complejidad de la función y en sus parámetros estadísticos (24).

La base matemática para el desarrollo de métodos de calibración multivariable aplicado a técnicas espectroscópicas de análisis instrumental, tiene su origen en el álgebra matricial. Donde “S” espectros de calibración son medidos a “W” discretas longitudes de ondas, es posible construir una matriz de datos espectrales “X” de dimensiones “W x S” que contiene un espectro en columna. De la misma manera es posible construir un vector “Y” de dimensiones “S x 1” que contiene los valores de referencia de las muestras de calibración. El objetivo de la calibración multivariable es calcular un vector “p” de dimensión “W x 1”, por lo que el objetivo es establecer un modelo de calibración que pueda ser aplicado al espectro de una muestra desconocida para estimar a “y”, lo cual equivale a calcular el vector de predicción “p” que se representa por

$$y = X^t p + e \quad (10)$$

Donde X^t es la transpuesta de la matriz “X” y “e” es un vector de dimensión “S x 1”, llamado vector error que corresponde a la diferencia entre los valores de referencia “y” y los valores estimados por el modelo “y”.

$$\hat{y} = X^t p \quad (11)$$

Generalmente el vector “p” se estima minimizando la suma de los cuadrados de los errores mediante la siguiente expresión:

$$e^t e = |e^2| = (y - X^t p) \quad (12)$$

La calibración multivariable nos permite cuantificar y mejorar la velocidad del análisis a partir de señales analíticas obtenidas de las muestras a las que se les predijo el valor de la propiedad estudiada mediante los modelos de calibración; ya que definen el rango sobre el cual las mediciones son válidas y demuestran si la linealidad y sensibilidad de los análisis son válidos para proveer estimaciones de

características físicas y químicas para los materiales a los cuales se les desarrolló la calibración (25).

4.3.1.1 Métodos de calibración multivariable. Entre los métodos de calibración multivariable que existen los más aceptados hasta el momento son los propuestos por Martens & Naes (26), basados en los siguientes aspectos:

Según la relación entre las variables dependientes e independientes: ésta relación puede ser descrita mediante un modelo lineal o no lineal.

Según la forma de encontrar la relación entre las variables: pueden ser métodos directos, donde los parámetros de calibración se calculan directamente a partir de la señal de cada uno de los analitos en forma individual, o métodos indirectos, donde tales parámetros se calculan a partir de las señales analíticas de las mezclas de componentes.

Según la variable que se defina como dependiente o independiente: si la calibración sigue criterio relacionado con la ley de Beer donde la señal analítica actúa como variable dependiente y la concentración como variable independiente se tiene un método de calibración directa. En caso contrario se tiene un método de calibración inversa.

4.3.1.2 Construcción de modelos de calibración multivariables. Las etapas que se requieren para la construcción de modelos de calibración multivariable a partir de mediciones espectrales sobre la muestra en estudio son las siguientes:

Hacer una selección de las muestras de manera que éstas sean representativas presentando la máxima variabilidad física y química esperada en ellas para las cuales será aplicado el modelo.

Realizar una caracterización de las muestras de calibración por un método de referencia previamente establecido, el cual haya sido evaluado estadísticamente y sea totalmente confiable.

Selección de las condiciones experimentales óptimas de adquisición espectral; esto con el fin de realizar un registro de la información analítica.

Realizar el pretratamiento a la señal espectral y se aplican técnicas de regresión sobre los datos espectrales con el objetivo de minimizar las contribuciones no deseadas en la señal analítica que puedan disminuir la reproducibilidad provocando además contribuciones erróneas de los parámetros deseados. Se realiza el pretratamiento de la señal comenzando con un suavizado que reduce el ruido de los datos seguido de una normalización con el objetivo de poner los datos aproximadamente en la misma escala, luego se hace una centralización que evita que ciertos puntos tengan más peso que otros en el modelo y por último la derivación de diferente orden que extrae información detallada que no puede ser observada en el espectro normal.

Dentro de las técnicas de correlación más comunes se tiene la regresión lineal múltiple (MLR), la regresión por componentes principales (PCR) y la regresión por mínimos cuadrados parciales (PLS).

Se valida el modelo de calibración comparando estadísticamente con los valores de referencia, los resultados obtenidos al aplicar el modelo generado sobre un grupo de muestras. El modelo es de calibración cruzada cuando son empleadas muestras incluidas en la calibración del modelo y se llama calibración externa cuando no se emplean muestras incluidas en la calibración del modelo, si no otras muestras diferentes.

La ecuación número 13 nos muestra el error medio de calibración que es un parámetro estadístico que se da mediante el cuadrado de las diferencias entre los valores de referencia y los predichos por el modelo de calibración.

$$RMSEC \sqrt{\frac{\sum_{i=1}^c (y_i - \hat{y}_l)^2}{n_c}} \quad (13)$$

Donde y_i es el valor de referencia de la respuesta, \hat{y}_l es el valor de la respuesta predicho por el modelo y n_c es el número total de muestras en el conjunto de calibración.

En el modelo de calibración cruzada como lo mencionamos anteriormente se realiza empleando los datos del conjunto de calibración para validar el modelo dejando una parte de la muestra fuera del conjunto de calibración y realizando el modelo con las muestras restantes prediciendo luego las descartadas esto quiere decir que primero se calcula el modelo dejando un grupo de cancelación fuera y se calcula la ecuación de la regresión con los demás datos; una vez realizada la calibración se pone a prueba el modelo en los datos del grupo de cancelación excluidos. Se repite el proceso de manera sucesiva, dejando cada vez un grupo de cancelación (grupo perteneciente al conjunto de calibración inicial) fuera del proceso del cálculo. Si el número de grupos de cancelación es igual al número de objetos en el conjunto de cancelación, cada vez quedará un objeto fuera del cálculo del modelo y el proceso se denomina validación cruzada total. Para validar un modelo de validación cruzada se debe definir la raíz cuadrada de la media del error al cuadrado (RMSECV) (27) (28) como:

$$RMSECV \sqrt{\frac{\sum_{i=1}^c (y_i - \hat{y}_{cv,i})^2}{n_c}} \quad (14)$$

En la última etapa entonces implementamos el modelo creado al análisis de muestras desconocidas y se instala el modelo como herramienta de análisis de rutina realizando periódicas revisiones para evaluar su desempeño (29).

4.3.2. Calibración de métodos basados en reducción de variables. Estos métodos tienen grandes ventajas ya que permiten cuantificar analitos individuales en mezclas sin la necesidad de conocer los otros componentes y además utilizan la información contenida en todo el espectro, y la "comprimen" en un pequeño número de variables. Esto evita la necesidad de seleccionar variables y facilita la detección de interferencias y los valores atípicos.

El número de variables de un sistema se puede reducir de varias maneras incluyendo el uso coeficientes de Fourier, entre otros. Sólo los dos más ampliamente utilizado para la calibración multivariable en química analítica son la regresión de componentes principales (PCR) y la regresión por mínimos cuadrados parciales (PLSR). Ambos siguen los principios de la regresión de componentes principales.

Tanto PCR y PLSR asumen que la información contenida en las variables medidas se puede concentrar en un número menor con el fin de reducir la relación señal/ruido sin perder la información pertinente. Los dos métodos de regresión simplifican la construcción del modelo de calibración y la interpretación de los resultados. Su importancia radica en la capacidad de resolver problemas, al menos en parte, típicos, como los siguientes:

- a) *una escasa selectividad*: mediciones de las variables X pueden ser influenciados por la presencia de interferentes que acompañan al analito. Esto requiere el uso de variables x para predecir y.
- b) *La colinealidad*: la información contenida en X puede ser redundante o incluso correlacionados.
- c) La falta de conocimientos precisa de la influencia de Y sobre X. Uno no puede saber todos los componentes de la muestra que influyen en X o sus interacciones mutuas. Además, la respuesta del instrumento puede ser imposible para linealizar en su totalidad.

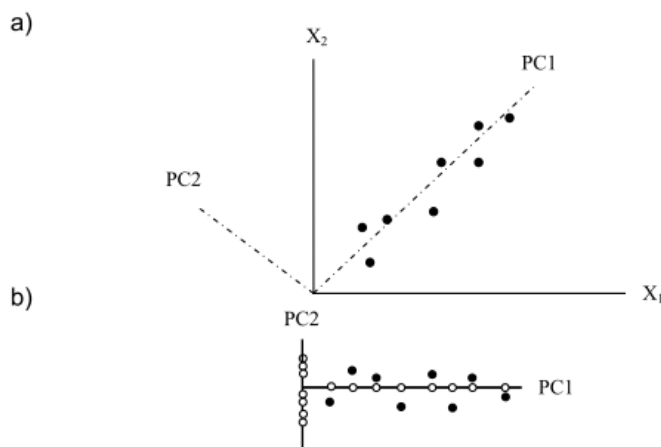
Además de las características típicas de métodos de reducción de variables, PCR y PLSR utilizar un espacio ortogonal para la regresión, lo que evita los problemas derivados de colinealidad entre las variables, (30). En pocas palabras, la regresión de componentes principales (PCR) implica un análisis de componentes principales (PCA) de la matriz de datos, X, seguido de regresión de mínimos cuadrados entre las puntuaciones de las muestras de calibración y seleccionar los valores de referencia del parámetro a ser modelado.

4.3.2.1. Análisis por componentes principales (PCA). Análisis de componentes principales es un método de compresión variable que reduce el conjunto de datos de la matriz X ($K \times N$) a un número mucho menor de variables denominadas componentes principales (PCs). Esto quiere decir que reduce el número de variables experimentales usando nuevas variables (los componentes principales) que son combinaciones lineales de las variables originales. El primer componente principal está dirigido en la dirección de máxima variabilidad de los datos, el segundo componente principal es ortogonal al primer componente principal y está dirigido en la segunda dirección de máxima variación de los datos, y así sucesivamente. La idea fundamental del análisis de componentes principales en su aplicación a datos espectroscópicos es reducir una matriz de datos experimentales al menor número posible de componentes abstractos cuya combinación lineal permita reproducir toda la variación sistemática que hay en los datos experimentales. Este número mínimo de componentes abstractos proporciona el número mínimo de componentes espectrales presentes en la muestra (31).

En la **figura 8a** podemos observar los métodos con dos variables y dos componentes principales; los componentes principales son representados en la figura mediante líneas de trazos sucesivos, formando trazos sucesivos unos con otros llamado ortogonalidad. En la **figura 8b** se observan los puntos referidos a estos nuevos ejes y la proyección de los puntos sobre PC_1 y PC_2 ; de esta forma se pueden trabajar en una dimensión (z_1) en lugar de trabajar en dos dimensiones (X_1 y X_2), ya que Z_1 recoge la mayor parte de la variación pudiendo de esta forma reducir la cantidad de datos con los que se está trabajando a Z_1 .

Podemos observar además que PC_1 se encuentra en la dirección de máxima variación, pero manteniendo el ángulo entre los ejes y PC_2 se encuentra en la dirección de la siguiente mayor variación continuando con esta secuencia, de manera que PCA es equivalente a una rotación de los ejes originales (32).

Figura 8. a) Representación de dos componentes principales, PC_1 y PC_2 , para dos variables, X_1 y X_2 . b) Proyección sobre su eje de los puntos referidos sobre PC_1 y PC_2

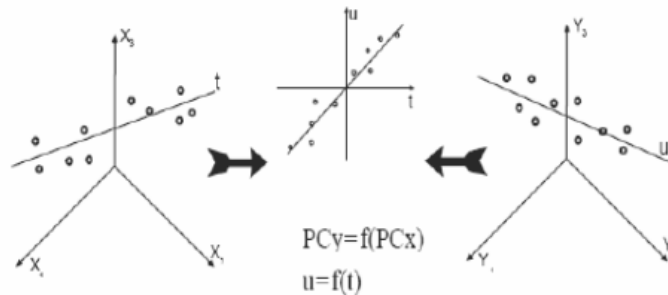


Fuente: MILLER N: James. Estadística y quimiometría para Química Analítica 2002.

4.3.2.2. Regresión por mínimos cuadrados parciales (regresión PLSR). El método PLSR, que fue presentado por Wold (1975), ha sido utilizada como una alternativa a la regresión de mínimos cuadrados ordinaria para resolver los problemas que implican alta colinealidad o la necesidad de calcular variables Y correlacionados (18). Es un método estadístico que está relacionado con la regresión de componentes principales ya que es la reducción de la dimensión en regresión múltiple, con la garantía de que las primeras componentes ortogonales mejoren la predicción. Como es sabido, esta es una característica que no poseen las componentes principales, lo que podría ser una de las ventajas de PLS sobre PCA. La estimación de las ponderaciones según el análisis de componentes principales es “óptima”, pero se pierden las relaciones de dependencia. En otras palabras podría decirse que es una generalización del ACP en el sentido de la introducción de relaciones (33).

En PLS se modela simultáneamente las matrices X y Y para encontrar un conjunto de variables latentes (VL) en X que mejor predicen las variables latentes en Y (figura 9).

Figura 9. Descripción gráfica del método de regresión PLS



Fuente: lesly viviana meléndez correa. *predicción del análisis sara de crudos colombianos*, 2010.

El PLS utiliza información contenida en la matriz de datos espectroscópicos (matriz x) en la etapa de calibración como la información contenida en la matriz de la propiedad a determinar (matriz y), para encontrar un conjunto de variables auxiliares llamadas variables latentes, VL, o factores en x que mejor predicen las variables latentes en Y (34).

Una vez centradas o auto escaladas las matrices X y Y escaladas, cada una de las matrices se descompone en una suma de A factores, calculando simultáneamente (35)

$$X = TP^t + E = \sum_{a=1}^A t_a p_a^t + E \quad (15)$$

$$Y = UQ^t + F = \sum_{a=1}^A u_a q_a^t + F \quad (16)$$

En donde la ecuación 15 es la descomposición de la matriz de datos espectroscópicos X, siendo T la matriz de Scores, P la matriz de *loadings* y E la matriz de residuales. Para la ecuación 16 que representa la descomposición de la matriz de la propiedad a determinar Y, ocurre lo mismo U es la matriz de Scores, Q la matriz de *loadings* y F la matriz de residuales.

La **ecuación 17** se muestra la relación interna entre los Scores de los bloques X y Y debido a que la descomposición de ambas matrices no es independiente si no que se realiza de forma simultánea y podemos observar además en la **ecuación 18** el coeficiente de regresión para cada uno de los factores PLS (b_a).

$$U_a = b_a t_a \quad (17)$$

$$b_a = \frac{u_a t_a}{t_a^T t_a} \quad (18)$$

La matriz Y puede ser calculada por medio de la ecuación 19 debido a que los coeficientes b_a hallados para cada componente son agrupados en una matriz diagonal B, que tiene los coeficientes de regresión de los Scores T y U de las matrices X y Y respectivamente.

$$Y = TBQ^T + F$$

Desde su formulación original, PLSR se ha asociado con otros métodos matemáticos y algoritmos. Los algoritmos más utilizados para implementar PLSR son mínimos cuadrados parciales iterativos no lineales (NIPALS) y SIMPLS. Entre los más recientes, sin embargo, se apartan del procedimiento iterativo clásica y facilitan la regresión más global y más rápida. En cualquier caso, las variables y están relacionados con las variables x a través de variables auxiliares llamadas variables latentes, o factores PLSR o componentes, que son combinación lineal de las variables x_1, x_2, x_k , y muy similar a la PCs calculados por PCA y utilizados para la PCR. Una descripción detallada de este método de calibración se puede encontrar en otra parte (36).

La diferencia de PCR es que PLSR tiene por objeto garantizar que las primeras variables latentes contendrán toda la información de uso predictivo posible. Para este propósito, el algoritmo PLSR utiliza la información contenida tanto en la matriz de datos espectroscópica, X, y la matriz de concentración, Y, durante la calibración y comprime los datos de tal manera que la varianza más tanto en X e Y

se explica. De esta manera, PLSR reduce el impacto potencial de las grandes, aunque irrelevantes, variaciones en X durante la calibración. En PLSR, cada componente se obtiene mediante la maximización de la covarianza entre Y y cada función lineal posible de X (18).

4.3.3. Validación del modelo. Una vez creado el modelo de calibración se debe validar empleando muestras independientes a las usadas en la calibración, esto con el fin de comprobar que no exista un error sistemático entre las predicciones que realiza el modelo y los valores proporcionados por el método de referencia; midiendo el grado de concordancia entre las predicciones del modelo y los valores del método de referencia ya que los métodos de calibración como PCR y PLS no se apoyan directamente en un modelo teórico y por lo tanto pueden incorporar variabilidad de los datos no necesariamente relacionada con la propiedad de interés por lo que deben ser validados.

4.3.4. Predicción de muestras desconocidas. Una vez el modelo es aceptado puede ser empleado para analizar otro tipo de muestras que cumplan con las condiciones de las muestras usadas para la calibración. En esta etapa se debe continuar con los test para detectar muestras discrepantes, outliers (capacidad de los métodos multivariantes de detectar la observación u observaciones inconsistentes con el resto de los datos), con el fin de detectar la presencia de extrapolaciones al modelo, presencia de nuevas interferencias, fallos instrumentales etc. En este caso se pueden utilizar medidas del *leverage* (medida de la influencia de una muestra en relación al modelo; entre más elevado el valor *leverage*, la muestra está muy alejada del centro del modelo (37) de las muestras, y del residual espectral. Herramientas para el control estadístico multivariante, como el estadístico T^2 de Hotelling (mide la variación de cada muestra dentro del modelo PCA y se calcula como la suma de los cuadrados de los Scores (38) y el estadístico Q , que se pueden utilizar también para la detección de outliers, ya que

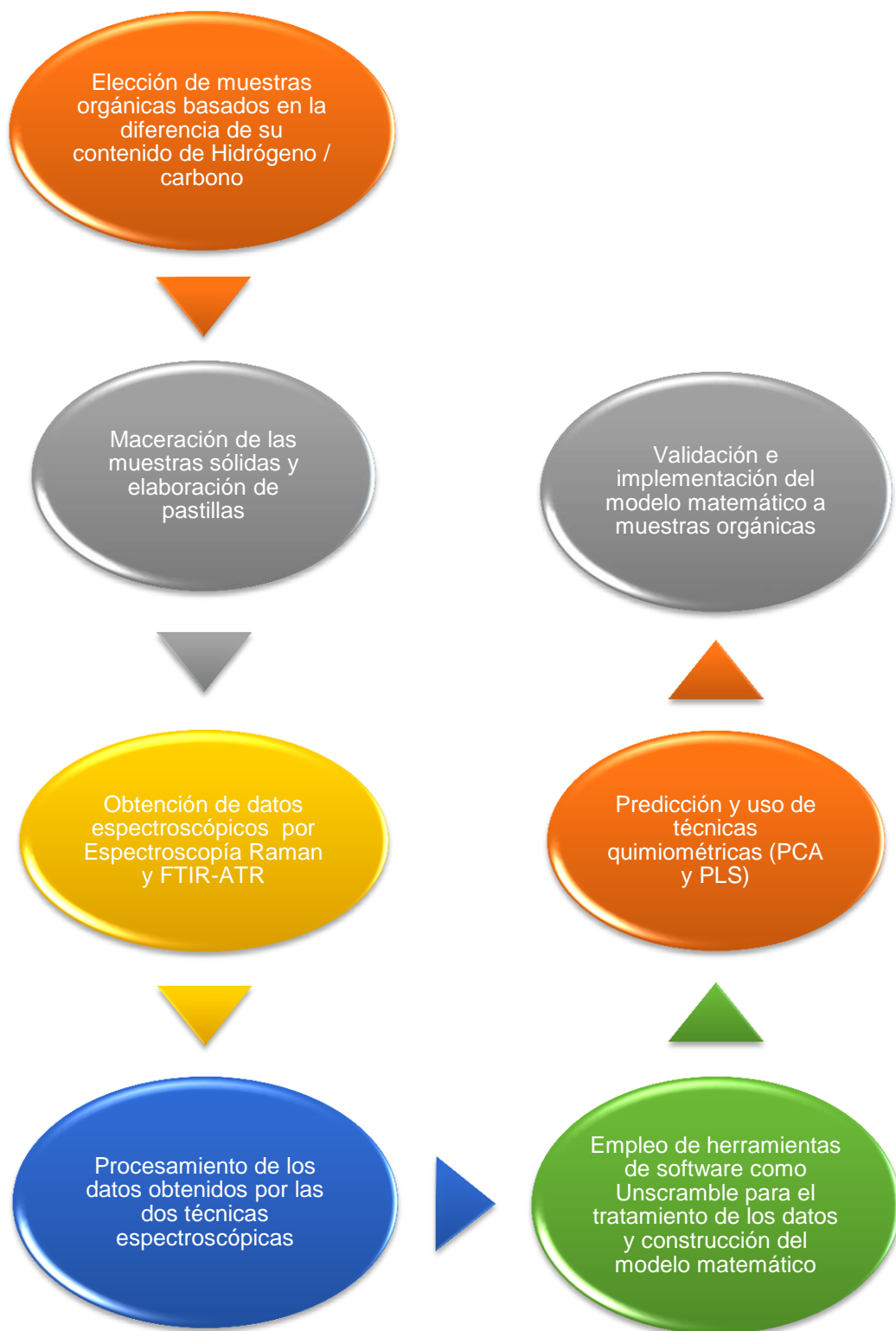
proporcionan una información similar al *leverage* ($EL T^2$) y al residual espectral (el estadístico Q). En esta etapa detectar outliers es de suma importancia puesto que la predicción de estas muestras puede diferir significativamente del valor verdadero.

5. METODOLOGÍA

En este trabajo investigativo se empleó técnicas espectroscópicas como espectroscopia Raman y espectroscopia infrarroja con reflectancia total atenuada (FTIR-ATR) aplicando métodos Quimiométricos y aprovechando el conocimiento, avances investigativos y los recursos con los que cuenta la Universidad Industrial de Santander; en la cual se han implementado modelos matemáticos para obtención de parámetros fisicoquímicos en crudos y otras sustancias y se desea acondicionar este método para establecer la relación hidrógeno vs carbono en muestras orgánicas.

A continuación se describe el procedimiento realizado para el desarrollo y obtención de dos modelos de predicción de la relación H/C de moléculas orgánicas a través de espectroscopia Raman, IR y PLS. Para establecer el procedimiento a seguir se tuvo en cuenta los trabajos realizados en el área de quimiometría y predicción de propiedades fisicoquímicas de crudos y fracciones realizados en el grupo de investigación de espectroscopia atómica y molecular LEAM de la Universidad Industrial de Santander (11) (29).

Figura 10. Esquema del desarrollo de la metodología



5.1 SELECCIÓN DE LAS MUESTRAS ORGÁNICAS

Las muestras orgánicas fueron proporcionadas por los laboratorios de síntesis orgánica, laboratorio de espectroscopia atómica y molecular (LEAM) y en la bodega de almacenamiento de reactivos para estudiantes de pregrado de química. Para la selección de las muestras se tuvo en cuenta el contenido de hidrógenos y carbonos presentes dividiendo el número total de hidrógenos por el número total de carbonos presentes en cada muestra y de esta forma obtener una relación H/C; para desarrollar un modelo del rango de calibración de 0,5 a 2,5 como se puede observar en la **tabla 3**.

Tabla 3. Relación H/C en muestras orgánicas usadas para el modelo de calibración

Muestras orgánicas	Fórmula química	Contenido de Hidrógenos	Contenido de Carbonos	Relación H/C
Antraquinona	$C_{14}H_8O_2$	8	14	0,57
Isatina	$C_8H_5NO_2$	5	8	0,62
Antraceno	$C_{14}H_{10}$	10	14	0,71
Naftaleno	$C_{10}H_8$	8	10	0,8
Difenilamina	$C_{12}H_{11}N$	11	12	0,92
Ácido antranílico	$C_7H_7NO_2$	7	7	1
Acetanilida	C_8H_9NO	9	7	1,12
Ácido Adípico	$C_6H_{10}O_4$	10	8	1,66
1,4 Dioxano	$C_4H_8O_2$	8	4	2
Dodecilhidrogeno sulfato	$CH_3(CH_2)_{11}OSO_3^-Na^+$	25	12	2,08
Isooctano	C_8H_{18}	18	8	2,25
n-Hexano	C_6H_{14}	14	6	2,33
Glicina	$C_2H_5NO_2$	5	2	2,5

No se realizó pre tratamiento a los compuestos orgánicos y a que no se requiere para las técnicas usadas en este estudio.

5.2 ESPECTROSCOPIA INFRARROJA CON REFLECTANCIA TOTAL ATENUADA (FTIR-ATR)

Los espectros MIR fueron tomados en el Laboratorio de Espectroscopía (LE) del parque tecnológico Guatiguará (PTG) sede Uis Guatiguará. Las muestras sólidas fueron previamente maceradas y se depositaron directamente en el porta muestras de la celda de ATR del equipo (**figura 11 B**) lo mismo con las líquidas de las cuales se dispuso de una a dos gotas en el equipo para su análisis. Se empleó el espectrómetro FTIR NICOLET iS50 (**figura 11 A**) que tiene las siguientes características:

- Celda ATR de cristal de diamante con reflexión simple de índice de refracción 2,4
- Fuente IR de larga duración Polaris™
- Accesorios especialmente diseñados y software integrado que lo hace especialmente útil para análisis de materiales de todo tipo.
- Cubre regiones del infrarrojo medio en el rango de 4000 a 400 cm^{-1}

Figura 11. A Espectrómetro FTIR NICOLET iS50 - B. Celda ATR del Espectrómetro



5.2.1 Adquisición de espectros MIR

Los espectros MIR fueron tomados en el laboratorio de espectroscopia de la Universidad Industrial de Santander y se tomaron bajo las siguientes condiciones:

Barridos: 32 Scan

Rango: 400 cm^{-1} a 4000 cm^{-1}

Resolución: 4 cm^{-1}

Ángulo de incidencia del haz: 45°

Velocidad del espejo del interferómetro: $0,4747\text{ cm}^{-1}/\text{s}$

Temperatura: 19°C

5.2.2 Análisis de espectros MIR de las 13 muestras orgánicas

Se tomó un espectro IR de las 13 muestras orgánicas (**figura 12**) en el cual se observan las bandas características de vibraciones de alargamiento C-H terciario cerca de la región 2890 cm^{-1} , las vibraciones de flexión en donde dos grupos metilo se encuentran unidos al mismo átomo de carbono como en el caso del compuesto 1,4 Dioxano y el Isooctano manifiestan una absorción en la región de flexión C – H en $1385 - 1380\text{ cm}^{-1}$ y 1370 a 1350 cm^{-1} aproximadamente, esto es muy común en hidrocarburos de cadena ramificada como el isooctano.

Las bandas más notables y de mayor información en el espectro son las que podemos observar en el intervalo de baja frecuencia de las regiones de 900 a 675 cm^{-1} pertenecientes a la flexión fuera del plano de los enlaces C–H de Hidrocarburos aromáticos mononucleares.

En el ácido adípico se presenta una banda de absorción de alargamiento C=O intensa en la región de $1870 - 1540\text{ cm}^{-1}$ características de la cetona, lo mismo que con la Isatina y ácido antranílico; su posición relativamente constante, su alta intensidad y relativa libertad de bandas de interferencia hacen de esta una de las bandas más fáciles de reconocer en el espectro infrarrojo.

Figura 11. Espectros Originales de las 13 muestras orgánicas

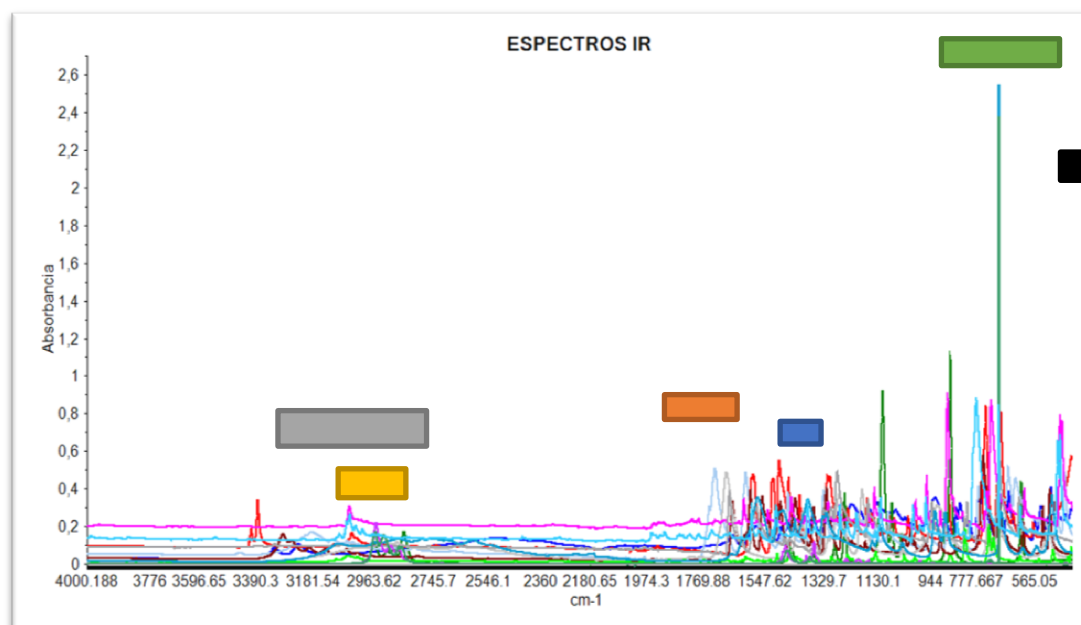


Figura 12. Espectros originales infrarrojo

- Vibraciones de alargamiento C-H terciario y vibración asimétrica de CH_3
- Vibraciones de flexión C – H
- Estiramiento y vibración de enlaces CH
- Derivados del Benceno (orto, meta, para, cis, trans y mono) sustituidos.
- Flexión fuera del plano de enlaces C-H
- Alargamiento C=O

5.3 ESPECTROSCOPIA RAMAN

Para la toma de datos en el espectrómetro Raman (**figura 13**) en muestras sólidas fue necesario macerarlas y elaborar pastillas para asegurar la uniformidad de la superficie impactada por el Láser, luego fueron depositadas en un porta objetos enfocando la sustancia con el microscopio (**figura 14 A**) el equipo se calibro previamente en la longitud de onda 732 nm con un porta objetos que contenía adherido una laminilla de sílice como se puede observar en **la figura 14.B**.

Para la toma de espectros de muestras líquidas, fue necesario depositar las sustancias en un micro tubo de cuarzo delgado (**figura 15B**) que permitió el traspaso del láser sobre las paredes del vidrio llegando a impactar la muestra contenida dentro del tubo que fue sellado con cera de abeja para evitar pérdidas por evaporación de las sustancias al ser calentadas por el rayo láser; evitando de esta manera además cualquier tipo de pérdida de muestra por derrame. Se analizaron 12 muestras orgánicas usando el montaje mostrado en la **figura 15**. Se empleó el espectrómetro Labram HR Evolution marca Horiba Scientific que tiene las siguientes características:

- Tres láseres de 478, 532 y 785nm.
- Rango de adquisición espectral de $50\text{-}4000\text{cm}^{-1}$
- Cuenta con dos rejillas de difracción de $600\text{ y }1800\text{mg/mm}$
- Tiempos de integración y número de Scan viables
- Microscopio cofocal con 4 objetivos de 10x, 50x y 100x.
- Detector cámara CCD

Figura 13. Espectrómetro Raman

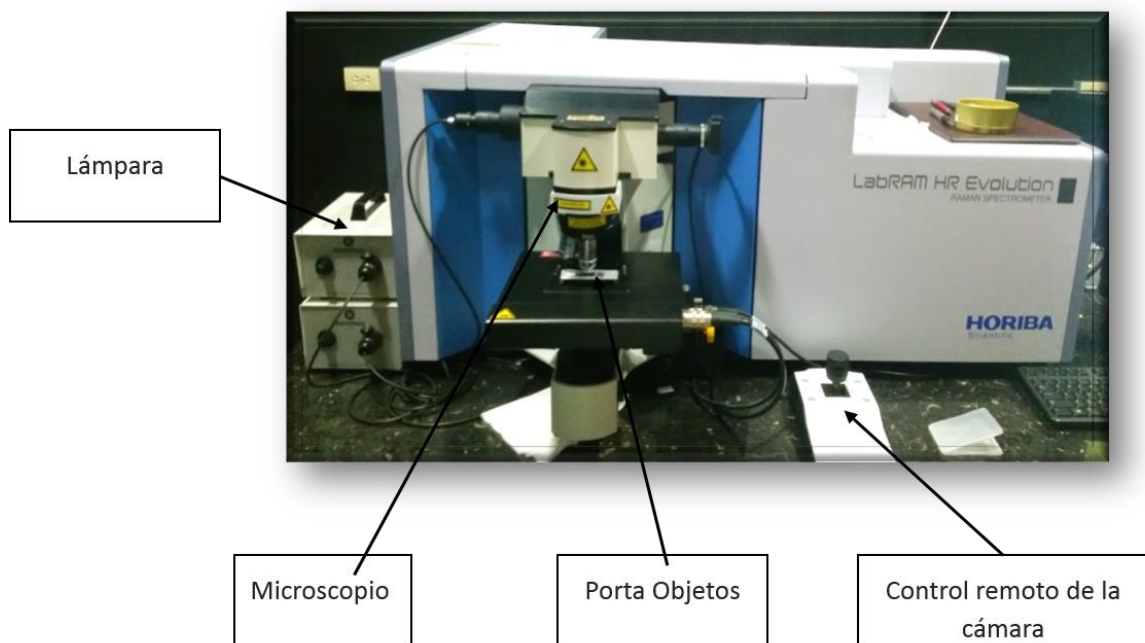


Figura 14. A. Microscopio - B. Porta objetos con laminilla de Sílice

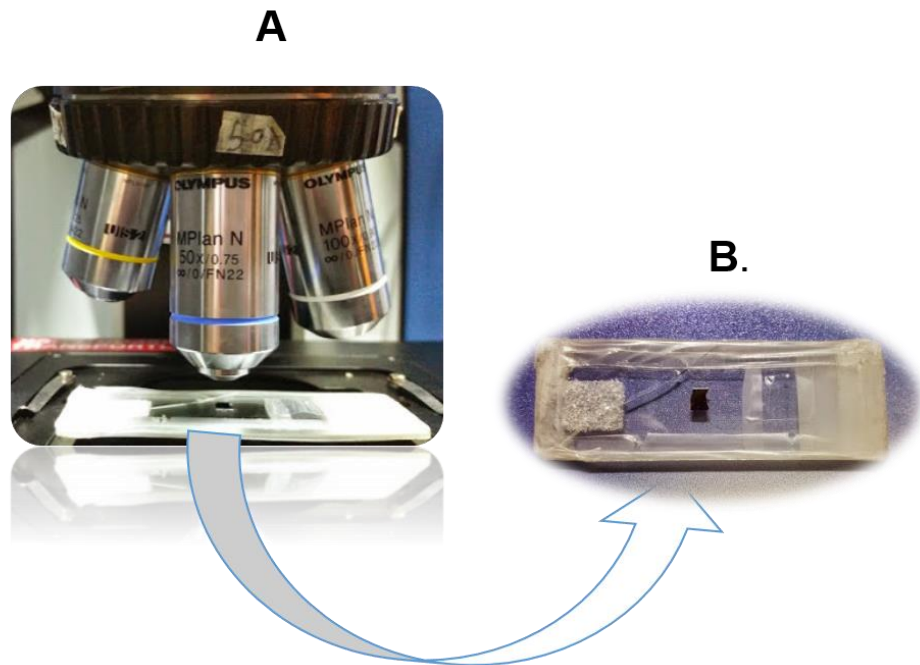
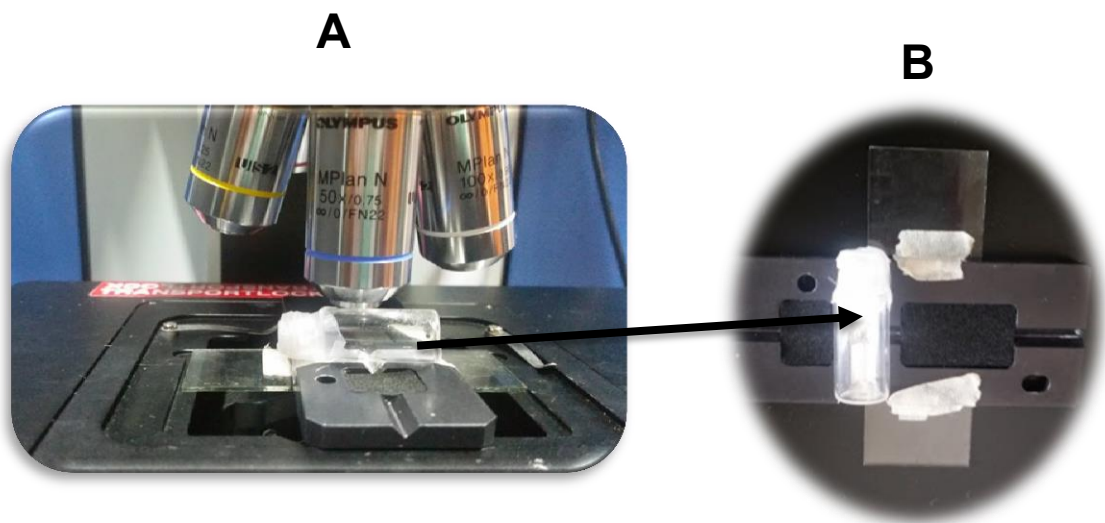


Figura 15. A. Montaje para la toma de espectros RAMAN de muestras líquidas - B. Vial con muestra orgánica líquida y tapón de cera de abeja.



5.3.1 Verificación del desempeño del Equipo. El desempeño del espectrómetro RAMAN fue medido realizando una prueba de repetibilidad como se puede observar en la **tabla 4** en donde se determinó el cálculo de las desviaciones estándar para las absorbancias medidas a 102,72

cm^{-1} 1749,63 cm^{-1} y 3156,43 cm^{-1} en 10 espectros del naftaleno encontrándose desviaciones estándar inferiores a 0,01 cm^{-1} asegurando de esta forma la repetibilidad de medición del espectrómetro.

Tabla 4. Prueba de repetibilidad del Espectrómetro RAMAN

ESPECTRO	LONGITUDES DE ONDA		
	102,72	1749,63	3156,43
1	8E ⁻⁰³	5,2E-05	2,4E-06
2	7 E ⁻⁰³	5,6E-05	4,0E-06
3	11 E ⁻⁰³	7,8E-05	3,7E-06
4	8 E ⁻⁰³	6,0E-05	3,7E-06
5	19 E ⁻⁰³	3,7E-05	2,6E-06
6	13 E ⁻⁰³	4,7E-05	3,2E-06
7	11 E ⁻⁰³	4,4E-05	4,3E-06
8	16 E ⁻⁰³	4,4E-05	2,9E-06
9	11 E ⁻⁰³	6,7E-05	8,3E-06
10	12 E ⁻⁰³	5,3E-05	4,5E-06
DESVIACIÓN ESTÁNDAR	4 E ⁻⁰³	1,2E-05	1,7E-06

5.3.2 Adquisición y Análisis de espectros RAMAN. Los espectros RAMAN fueron tomados en el Laboratorio de espectroscopia de la Universidad Industrial de Santander bajo los siguientes parámetros:

Hole= 100

Tiempo de adquisición= 10 s

Tiempo de acumulación= 10 acumulaciones.

Rango= 100 a 3200 cm^{-1}

Objetivo del microscopio= 50x

Grating= 600 gr/mm

Láser usado= 785nm

Filtro del láser = 100%

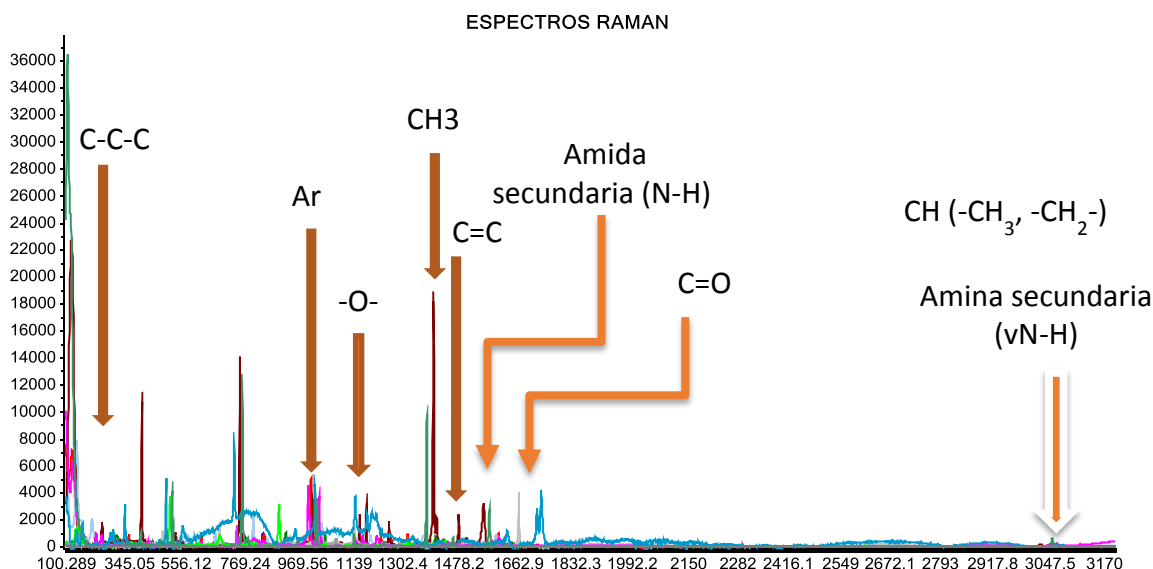
5.3.3 Análisis de los espectros obtenidos. La **figura 16** presenta el espectro Raman de 12 muestras orgánicas en las que se observan bandas en la región de 100 a 3200 cm^{-1} . Se observa una pequeña banda en 3055,89 cm^{-1} característico de los grupos CH (-CH₃, -CH₂-) ya que en espectroscopia Raman estos grupos pueden ser muy débiles o estar ausentes en el espectro; caso contrario se produce en espectroscopia Infrarroja en donde estas señales aparecen con mayor intensidad. En la región comprendida entre 3300 cm^{-1} a 1454 cm^{-1} aparece la señal característica de la vibración de enlaces nitrógeno-hidrógeno en los grupos amínicos (simétricos y asimétricos), aminas primarias y secundarias como en el caso de la difenilamina y la glicina.

Las bandas CH simétricas, asimétricas, alifáticas o aromáticas normalmente suelen aparecer entre los 2500 y 3000 cm^{-1} ; además se aprecia en 1413.09 cm^{-1} un pico característico de la banda CH₃ la cual se encuentra reportada en la literatura en la región comprendida entre 1400 a 1470 cm^{-1} . En la literatura se reporta que en la región de 1624 cm^{-1} la banda característica de la vibración de enlaces nitrógeno-hidrógeno en los grupos amínicos (simétricos y asimétricos)

En 1513 cm^{-1} se observa los grupos C=C que aparece alrededor de 1510. En la región de 1610 a 1550 cm^{-1} se puede observar una banda característica de la amida secundaria (δNH) que contienen las muestras Acetanilida y Satina y en la

región de 1700 a 1800 cm^{-1} se evidencia un pico prominente debido al enlace C=O. Los grupos aromáticos (Ar) suelen aparecer alrededor de 1000 cm^{-1} y en la **figura 16** de espectros originales Raman se puede observar que aparecen en la región de 994,9 cm^{-1} hasta 1036,9 cm^{-1} , también se presenta la vibración asimétrica del -O- en 1145 cm^{-1} que generalmente se encuentra entre el intervalo de 1060 a 1150 cm^{-1} característica del compuesto 1,4 Dioxano es de menor intensidad en Raman pero se presentan de mayor intensidad en Infrarrojo, esto debido a que el ángulo entre enlaces involucrados en el acoplamiento es mayor de 90°

Figura 16. Espectros Raman Originales de las 12 muestras orgánicas



5.4 PROCESAMIENTO DE DATOS ESPECTROSCÓPICOS

Los datos espectroscópicos obtenidos fueron analizados en el software The Unscramble versión X 10.3 suministrado por el Laboratorio de Espectroscopia

Atómica y Molecular (LEAM) de la Universidad Industrial de Santander quien cuenta con la licencia respectiva; y se hizo uso de él para el tratamiento de espectros haciendo correcciones de la línea base, ruido, diferencia entre las escalas etc.

Para procesar los datos como primera medida se realizó un análisis exploratorio de los datos espectrales, por lo cual se tomaron tres puntos de todo el espectro; cada punto representa una longitud de onda diferente en este caso representan las longitudes de onda $3048,4 \text{ cm}^{-1}$, $1661,8 \text{ cm}^{-1}$ y $688,4 \text{ cm}^{-1}$ para FTIR ATR, y $102,6 \text{ cm}^{-1}$, $1749,6 \text{ cm}^{-1}$ y $3156,3 \text{ cm}^{-1}$ para espectroscopia Raman con los datos espectroscópicos de las muestras orgánicas para cada longitud de onda se halló la desviación estándar.

Debido a que el software Unscramble presenta diferentes opciones para el pretratamiento de los datos, se realizaron seis normalizaciones contenidas en el software para el espectro completo y se calculó para cada una de la señales mencionadas la desviación estándar, encontrándose que el pretratamiento que mejor hace corrección de los espectros y que se ajustó satisfactoriamente a los datos fue la normalización por áreas para las dos técnicas analíticas.

5.4.1 Espectroscopía infrarroja. En la **tabla 5** se presentan las desviaciones estándar relativas obtenidas para los diferentes pretratamientos realizados a los 13 espectros de las muestras orgánicas; como se puede observar el que presenta menor desviación es la normalización por áreas y por lo tanto fue esta normalización la escogida.

La desviación estándar relativa (RSD) es una medida de dispersión de un conjunto de datos alrededor de su valor medio y es de suma importancia en el tratamiento de los datos ya que mediante el cálculo de la RSD podemos identificar los

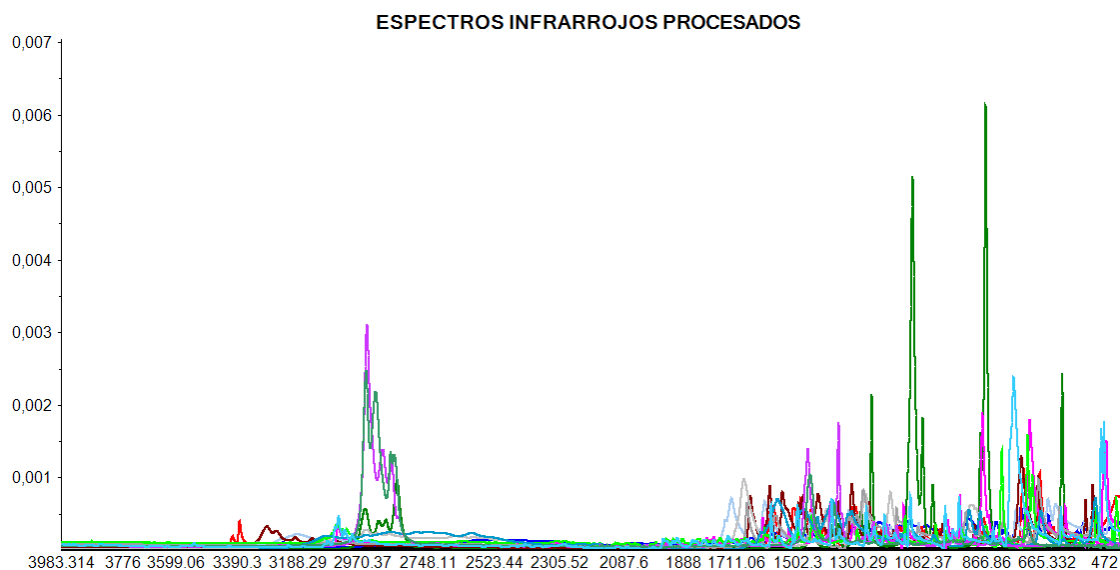
pretratamiento adecuados para aplicar a datos experimentales observando el menor valor obtenido de desviación estándar ya que esto representa menor dispersión, menor ruido y menor representatividad de la media.

Tabla 5. Desviación estándar para cada normalización con respecto a las longitudes de onda original en los espectros obtenidos por Espectroscopia Infrarroja

LONGITUD DE ONDA (cm^{-1})	3048,47	1661,88	688,47
<i>Datos originales</i>	0,096	0,12	0,22
<i>Normalización por Rangos</i>	0,078	0,21	0,33
<i>Normalización por vector Unitario</i>	0,007	0,01	0,019
<i>Normalización por Medias (NM)</i>	1,02	1,72	2,29
<i>Normalización Máxima</i>	0,078	0,21	0,33
<i>Normalización por picos</i>	8,98	10,79	27,95
<i>Normalización por Área (NA)</i>	0,0001	0,0002	0,0003

En la **figura 17** podemos observar los cambios ocurridos en los espectros infrarrojos cuando éstos son normalizados.

Figura 17. Espectros IR Normalizados



No se hizo mención al suavizado espectral de los espectros debido a que el software IR solution hace este pretratamiento automáticamente.

5.4.2 Espectroscopía Raman. Las longitudes de onda escogidas para el cálculo de la desviación estándar fueron: $102,6 \text{ cm}^{-1}$, $1749,6 \text{ cm}^{-1}$, $3156,3 \text{ cm}^{-1}$ con el objetivo de identificar el mejor pretratamiento para los datos de espectros Raman, al igual que para infrarrojo, la normalización por áreas, generó la menor desviación estándar.

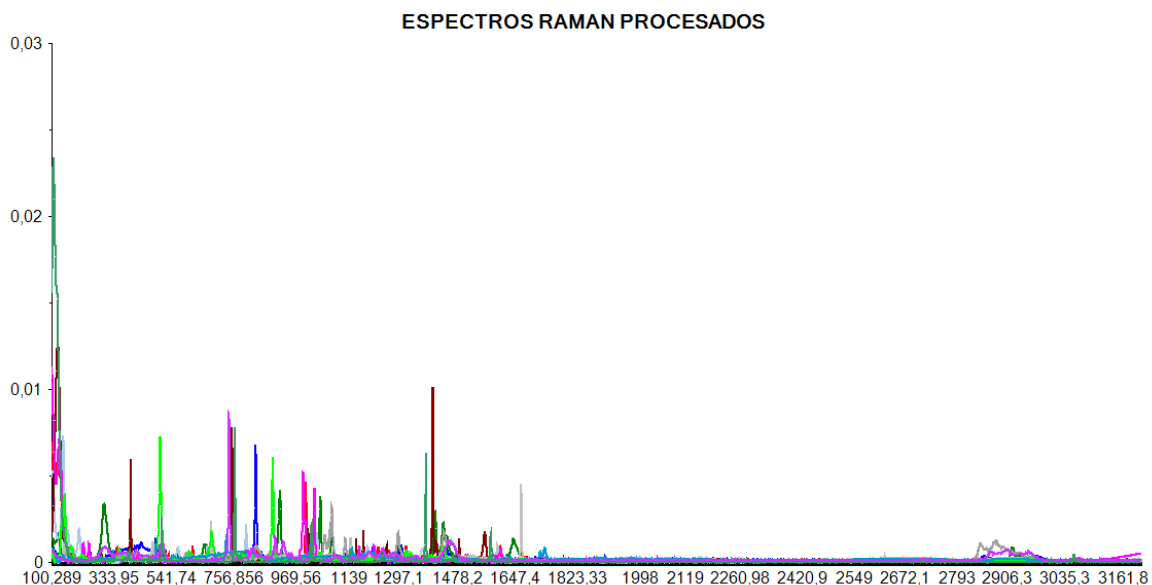
En la siguiente tabla se representan las RSD obtenidas para los diferentes pretratamientos realizados a los espectros Raman de las muestras orgánicas.

Tabla 6. Desviación estándar para cada normalización con respecto a las longitudes de onda original en los espectros obtenidos por Espectroscopia Raman

<i>LONGITUD DE ONDA (cm⁻¹)</i>	102,625	1749,6	3156,3
<i>Datos originales</i>	14017,7	5148,3	25,1
<i>Normalización por vector Unitario</i>	0,063	0,006	0,0004
<i>Normalización por Medias (NM)</i>	26,39	0,49	0,07
<i>Normalización Máxima</i>	0,29	0,16	0,003
<i>Normalización por Rangos</i>	0,29	0,16	0,003
<i>Normalización por picos</i>	0,08	0,56	0,007
<i>Normalización por Área (NA)</i>	0,005	0.00009	0,000013

En la figura 18 se muestran los espectros raman normalizados y con corrección de línea base.

Figura 18. Espectros Raman Normalizados



6. ANÁLISIS DE RESULTADOS

Después de realizar el pre-tratamiento a los espectros obtenidos por Raman e Infrarrojo; el siguiente paso fue la realización de los modelos de predicción que se describen a continuación.

6.1 ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)

Para la reducción del número de variables perdiendo la menor cantidad de información posible se realizó una descomposición de componentes principales del conjunto total de espectros corregidos por el pretratamiento realizado a las muestras orgánicas.

Los datos de varianza explicada son muy importantes para saber cuántos componentes principales vamos a utilizar en nuestro análisis. No hay una regla definida sobre el número que se debe utilizar, con lo cual se decide en función del número de variables iniciales (se busca reducirlas en la medida de lo posible) y de la proporción de varianza explicada acumulada.

6.1.1 Espectroscopia Infrarroja (FTIR-ATR). Para reducir la dimensionalidad del conjunto de datos, sintetizar la información, hallar las causas de la variabilidad e identificar las muestras que están correlacionadas entre sí, como primera medida se realizó una descomposición en componentes principales del conjunto total de espectros corregidos por el pretratamiento realizado.

En la **tabla 7** se puede observar la representación de los nueve primeros componentes principales (PC_1 a PC_9) en donde se muestra la varianza explicada por cada uno de ellos.

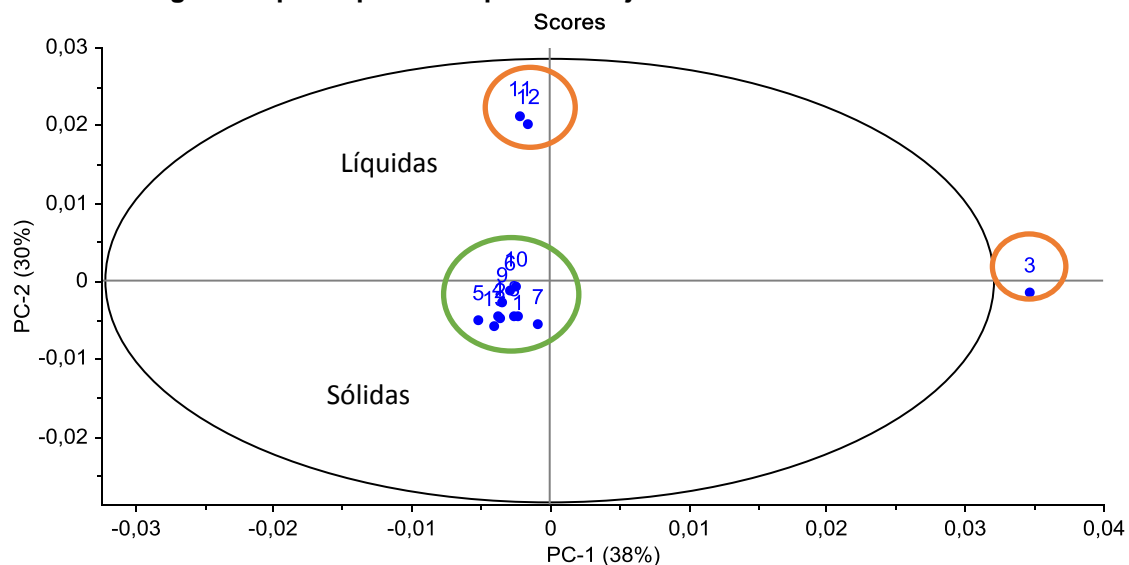
Tabla 7. Varianza explicada por los componentes principales de las muestras orgánicas por FTIR-ATR

Componente Principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
<i>PC₁</i>	38,02	38,02
<i>PC₂</i>	29,89	67,92
<i>PC₃</i>	8,31	76,22
<i>PC₄</i>	6,02	82,25
<i>PC₅</i>	4,22	86,47
<i>PC₆</i>	2,72	89,19
<i>PC₇</i>	2,58	91,77
<i>PC₈</i>	2,25	94,02
<i>PC₉</i>	1,95	95,97

Para obtener la mayor varianza explicada fue necesario el uso de 7 componentes principales que explican el 91,76 % de la misma, a partir del octavo componente el aumento de la varianza explicada acumulada es menor a 2,25 % por lo tanto se tomaron siete para el análisis por PCA.

El análisis por componentes principales genera un gráfico denominado Score que permite realizar una comparación bidimensional de dos componentes y de esta manera determinar las diferencias o semejanzas que existen entre las muestras e identificar aquellas atípicas o Outliers. (**Figura 19**).

Figura 19. Gráfico de Scores de los primeros componentes principales (PC1 vs PC2) de las 13 muestras orgánicas por espectroscopia infrarroja



En el gráfico de scores podemos apreciar números que representan cada muestra orgánica y se especifica en la tabla 8.

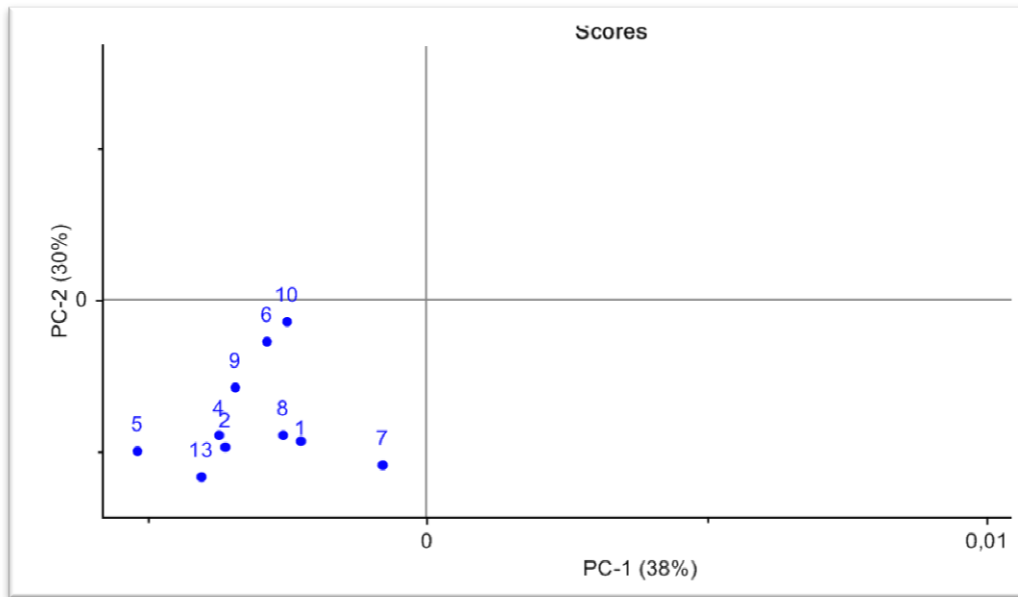
Tabla 8. Número asignado para cada muestra orgánica representados en el gráfico de Scores para espectroscopia infrarroja

<i>compuestos</i>	Numero asignado
<i>Ácido antranílico</i>	1
<i>Difenilamina</i>	2
<i>1,4 Dioxano</i>	3
<i>Isatina</i>	4
<i>Acetanilida</i>	5
<i>Ácido adipico</i>	6
<i>Antraceno</i>	7
<i>Antraquinona</i>	8
<i>Dodecilhidrogenosulfato</i>	9

<i>Glicina</i>	10
<i>Isooctano</i>	11
<i>Hexano</i>	12
<i>Naftaleno</i>	13

De acuerdo a la **figura 19**, se puede observar que todas las muestras se acumulan en los valores centrales de ambos componentes, permitiendo establecer que existe relación entre ellas según su ubicación en el plano $PC_1 - PC_2$, ya que cuanto más cerca están las muestras al eje de coordenadas del gráfico, más similares son con respecto a los dos componentes en cuestión. Dependiendo de la forma como se distribuyen los puntos en el gráfico de scores podemos inferir que las muestras se agrupan en dos formas; dependiendo de su estado físico, debido a que el grupo que representa la mayoría de las muestras están en estado sólido a temperatura ambiente (agrupación en verde) mientras el otro grupo en el que se encuentran las muestras 11,12 y 3 se encuentran en estado líquido (agrupación naranja). La muestra número 3 se encuentra alejada de las demás debido a que tiene variación en cuanto a su estructura molecular, ya que además de tener hidrógenos y carbonos también posee dos átomos de oxígeno ocasionando un cambio en sus modos vibracionales de tensión y flexión esto hace que sea un potencial Outliers (puntos fuera de la elipse o muestras atípicas) debido a que se ubicó fuera de los dos grupos establecidos en el gráfico de Scores. Las muestras 11 y 12 sólo contienen átomos de hidrógeno y carbono en su estructura y son las de mayor relación H/C de la mayoría de compuestos orgánicos exceptuando al componente Glicina (muestra número 10) que se encuentra en el grupo de los compuestos en estado sólido y está más próxima al primer componente principal (PC_1) como se puede observar en la **figura 20**.

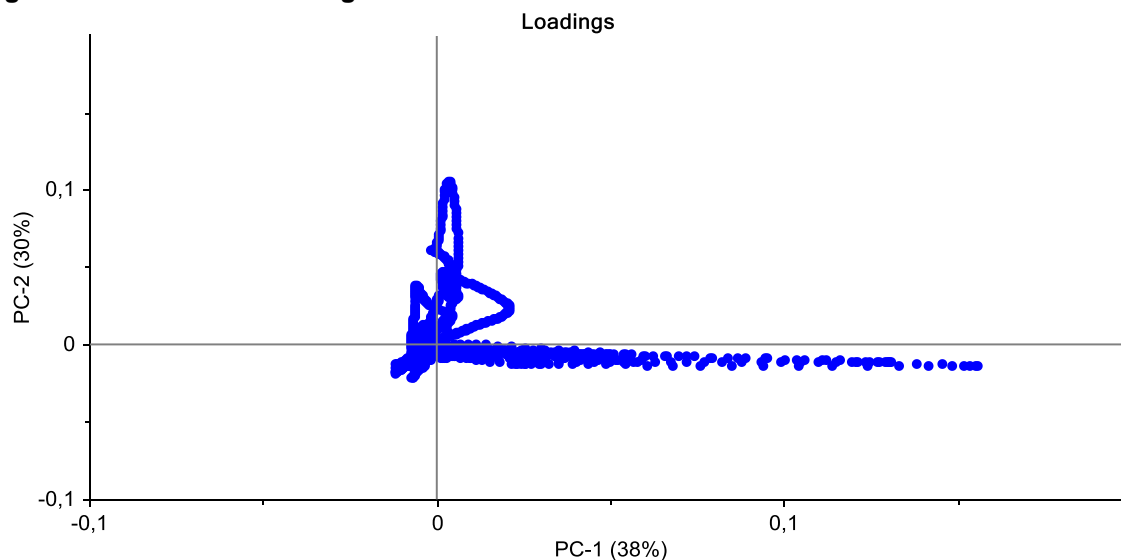
Figura 20. Gráfico de Scores de los primeros componentes principales (PC₁ vs PC₂) de las 10 muestras orgánicas sólidas a temperatura ambiente



La representación de los dos primeros componentes principales (PC₁ y PC₂) en el gráfico de Scores es una dispersión en dos dimensiones en los cuales se recoge el 28,02% y 29,89% de varianza respectivamente, lo que indica que entre los dos componentes se explica el 57,91% de la varianza total de las intensidades espectrales.

Haciendo una comparación del gráfico de Scores con el de Loadings representado en la **figura 21** se puede determinar las variables relevantes o influyentes en la distribución o cercanía de las muestras, ya que en los valores positivos para el eje PC₁ se encuentra las variables del espectro influyentes en las muestras líquidas 11 y 12 y en los valores negativos PC₁ se encuentra las variables x del espectro importantes de las muestras sólidas 1,2,4,5,6,7,8,9 y 10.

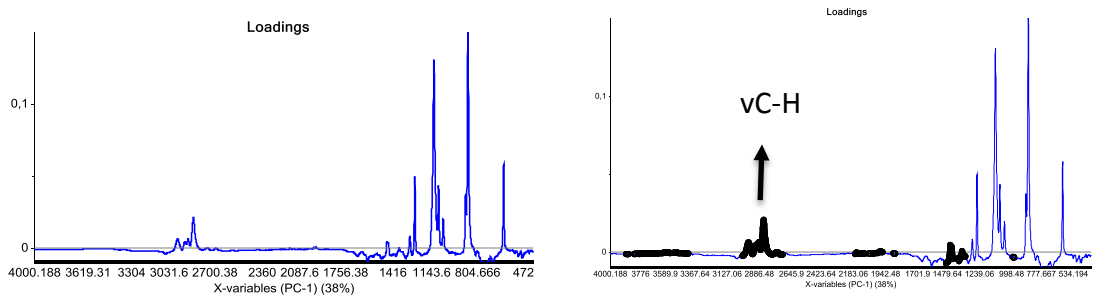
Figura 21. Gráfico de Loadings 2D



Representando el gráfico de Loadings de forma lineal (**figura 22**) podemos apreciar en negrilla las señales más intensas de las muestras ubicadas en la parte positiva del gráfico del Score (PC_2) mostrado en la **figura 19** que en este caso es el comportamiento lineal de las muestras líquidas Isooctano y Hexano quienes poseen únicamente átomos de hidrógeno y carbono en su estructura. En los puntos no sombreados del espectro se encuentran las señales de las demás muestras orgánicas representadas en el eje negativo de PC_2 en el gráfico de Score.

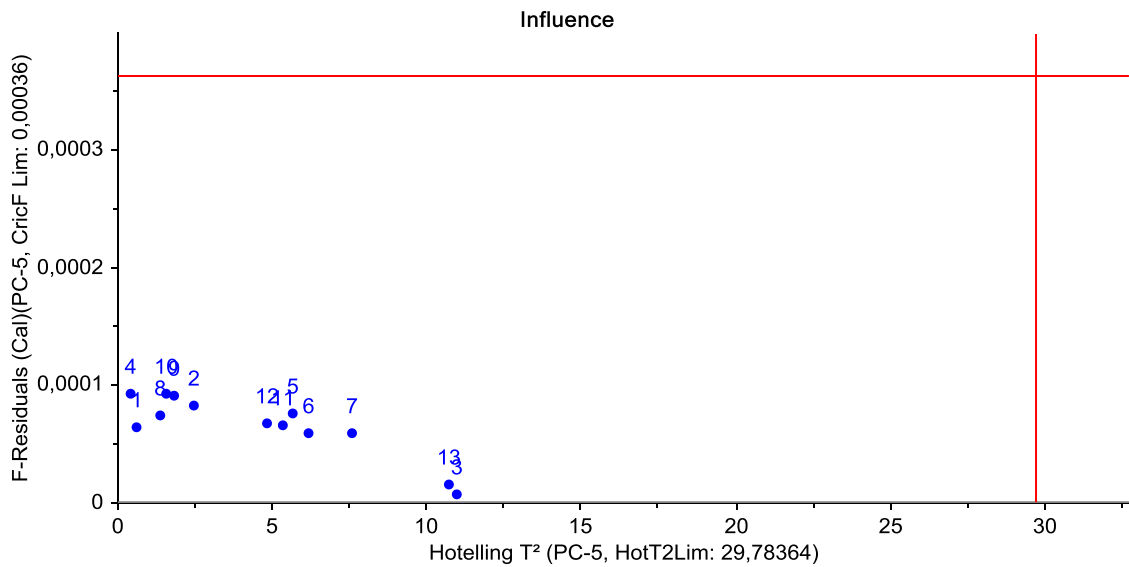
Las vibraciones simétricas y Asimétricas de CH alifático se presentan en los picos en negrilla de la región de $3029,671\text{cm}^{-1}$ a 2760cm^{-1} del gráfico de Loadings lineal mostrado a continuación.

Figura 22. Gráfico de Loadings lineal



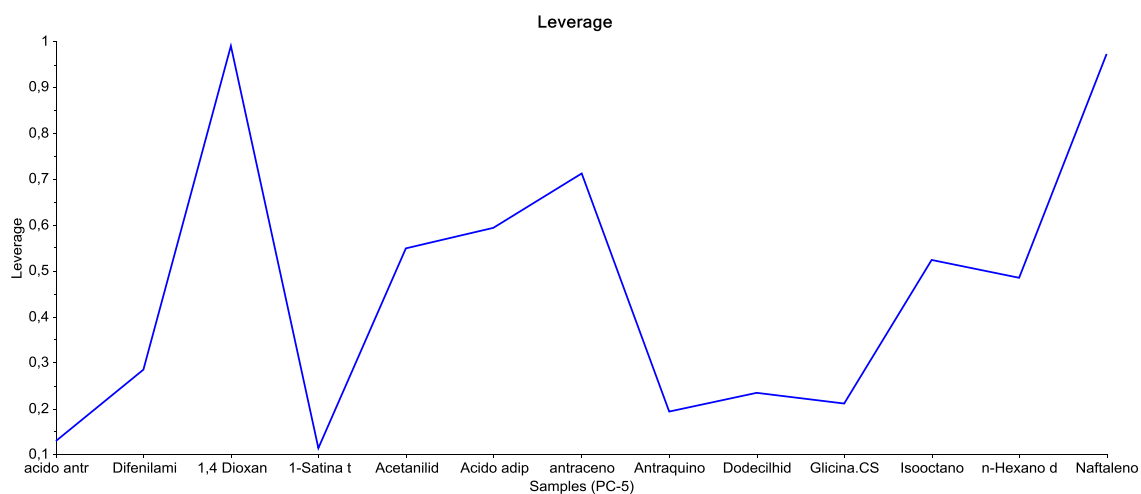
Para determinar valores atípicos u Outliers que son peligrosos e influyen en las muestras orgánicas se realizó el gráfico de la **figura 23**. No se observan muestras con alta varianza residual, ya que no se observan puntos desplazadas hacia la parte superior de la gráfica, por lo tanto podemos asegurar que las variables están bien representadas por el modelo que describe correctamente la mayoría de las muestras las cuales se pueden diferenciar mediante las características más comunes entre ellas.

Figura 23. Gráfico de Influencia para PC-5 para las 13 muestras orgánicas



Los puntos número 13 y 3 presentan mayor Leverage lo cual indica que pueden ser muy influyentes en el modelo de predicción debido a que se encuentran desplazados hacia la derecha de la **figura 23**. Como se evidenció anteriormente con el gráfico de Scores y el estadístico T^2 de Hotelling; el punto número 3 que representa la muestra orgánica 1,4 Dioxano se define como dangerous outliers, la cual puede distorsionar el modelo PLS con el fin de estar mejor descrito, lo que significa que el modelo tenderá a centrarse en la diferencia entre esta muestra y las demás, en lugar de describir más características generales comunes a todas, lo anteriormente se puede comprobar mediante el gráfico de Leverage de la **figura 24** en donde los puntos 13 (naftaleno) y 3 (1,4 Dioxano) presentan un pico más alto que las demás muestras orgánicas.

Figura 24. Gráfico de Leverage para PC5 en el modelo PCA



Como se puede ver en la figura anterior, no se presentan muestras con carácter negativo y no se observan muestras que sobrepasen el límite de influencia.

6.1.2 Espectroscopia Raman. En la descomposición de componentes principales se encontró que los primeros siete componentes ($PC_1, PC_2, PC_3, PC_4, PC_5, PC_6, PC_7$)

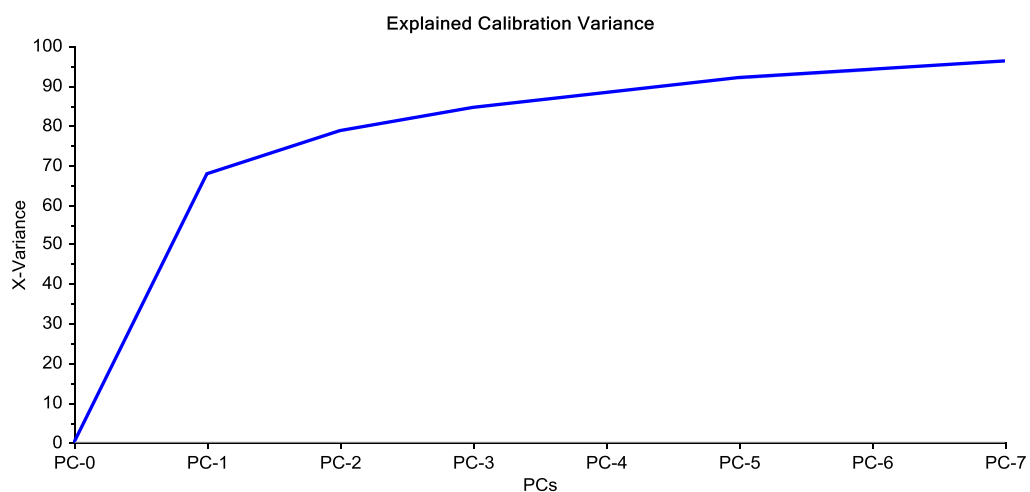
describen el 96% de la varianza explicada (tabla 9), después del sexto componente se explica menos del 2% de varianza.

Tabla 9. Varianza explicada por los componentes principales de las muestras orgánicas por Raman

Componente Principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
<i>PC₁</i>	67,90	67,90
<i>PC₂</i>	10,74	78,65
<i>PC₃</i>	5,85	84,49
<i>PC₄</i>	3,95	88,45
<i>PC₅</i>	3,49	91,93
<i>PC₆</i>	2,25	94,18
<i>PC₇</i>	1,97	96,16

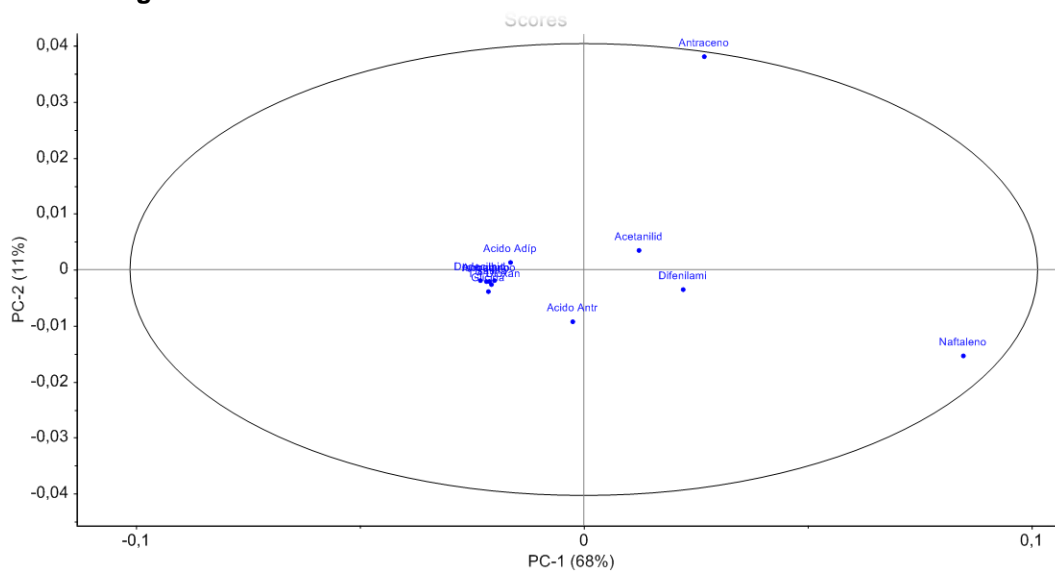
Como se observa en la **figura 25**, el modelo de predicción explica el 96 % hasta el componente principal 7 por lo que la gráfica continúa creciendo hasta este valor debido a la variabilidad de las muestras orgánicas en cuanto a su composición y estructura. Se nota un aumento significativo hacia *PC₁* debido a que éste componente explica el 67,9 % en la variabilidad de los datos (variable Y).

Figura 25. Gráfico de Varianza Explicada en el modelo de PLS para muestras orgánicas



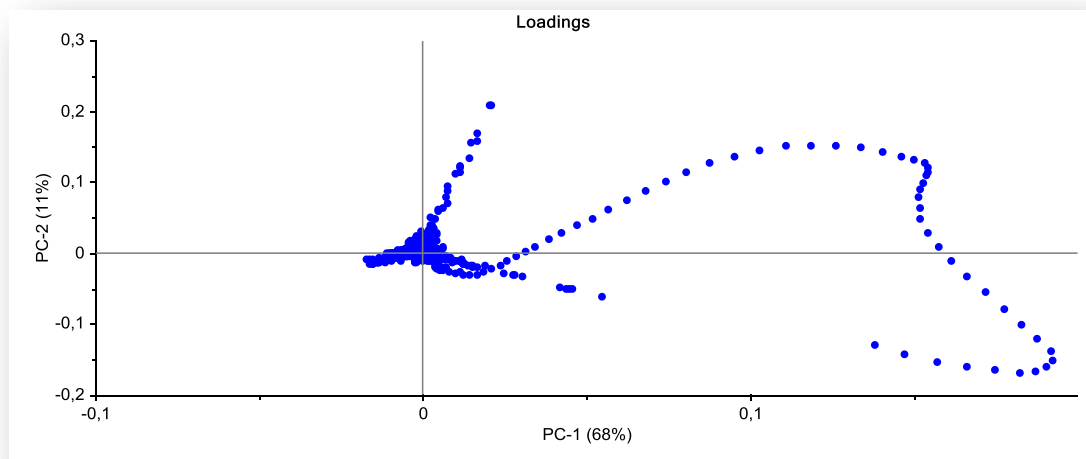
En el gráfico de Score representado en la **figura 26** para los dos primeros componentes principales (PC_1 y PC_2) se puede observar que no se encontraron agrupaciones naturales para las muestras ya que sus heteroátomos son muy diferentes entre una muestra y otra, además no se identificaron Outliers ya que ningún punto se salió de la elipse por lo que todas las muestras se encuentran bien representadas.

Figura 26. Gráfico de Scores de los componentes principales (PC_1 vs PC_2) de las 12 muestras orgánicas



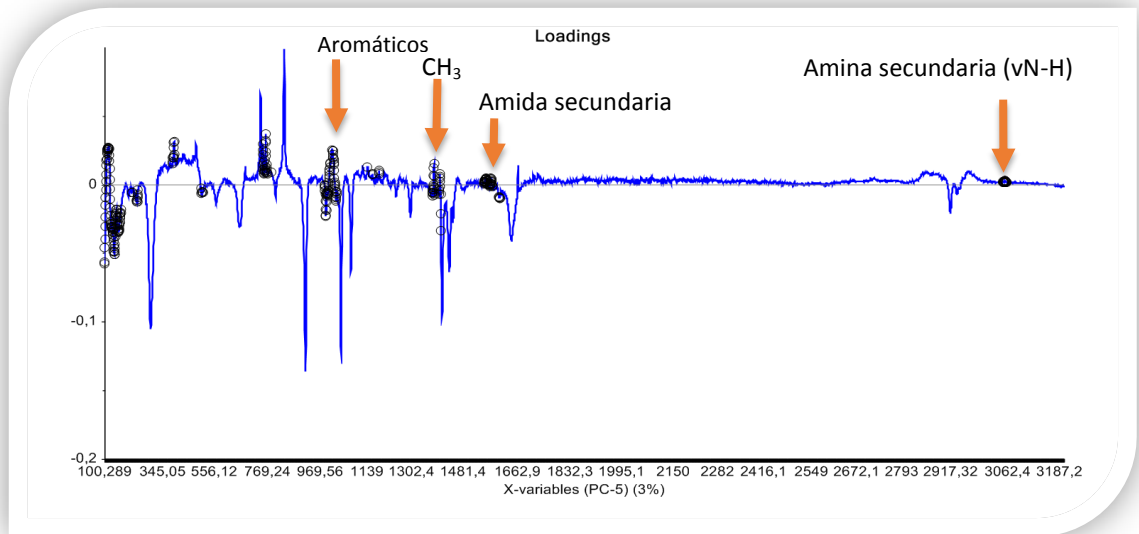
En el gráfico de Loadings 2D mostrado en la **figura 27** podemos observar las longitudes de onda más importantes para las mayores variaciones de los datos espectroscópicos. El eje positivo de PC₂ corresponde a las variables de las muestras orgánicas antraceno, acetanilida, difenilamina y naftaleno las cuales están influenciando en las muestras del segundo y cuarto cuadrante del gráfico de Score y se puede apreciar una característica común entre todas ellas y es que tienen anillos aromáticos unidos entre ellos como en caso del antraceno y naftaleno o el anillo aromático va unido únicamente a un átomo de nitrógeno en lugar de encontrarse unido a un átomo de carbono como en el caso de la Difenilamina y la Acetanilida propiedad que las hace diferentes al resto de compuestos que no tienen anillos aromáticos o en el caso de tenerlos están unidos a átomos de carbono.

Figura 27. Gráfico de Loadings 2D de 12 muestras orgánicas por Espectroscopia Raman



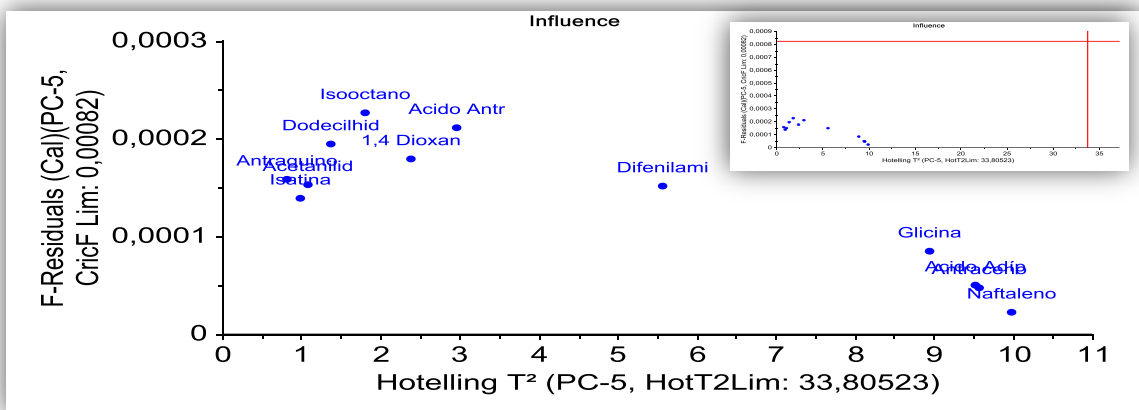
En la **figura 28** del gráfico de Loadings lineal se observan las señales más importantes para las muestras antraceno, acetanilida, difenilamina y naftaleno resaltadas en negrilla que se encuentran en el gráfico de Score para PC₁ positivo (**figura 26**).

Figura 28. Gráfico de Loadings Lineal



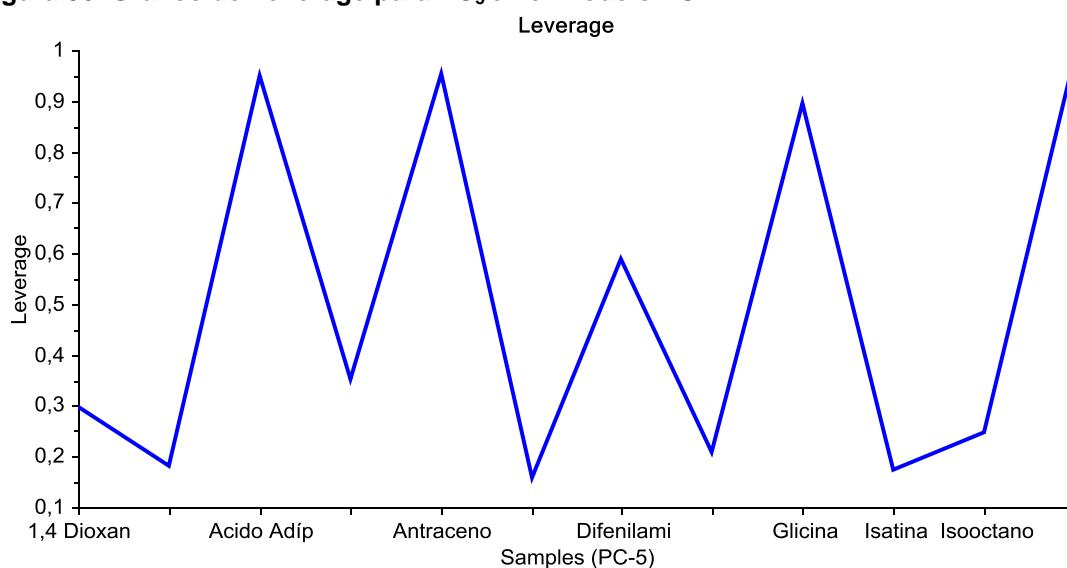
Como se evidencia en la **figura 29** no existen muestras que presenten una alta varianza residual, lo que indica que las variables de las muestras están bien descritas por los componentes, dentro del recuadro se puede observar que no existen valores atípicos, influyentes y peligrosos de las muestras analizadas por lo tanto se puede decir que los PCs describen las características más generales comunes a todas las muestras. El naftaleno, ácido adípico, antraceno y glicina que se encuentran levemente más inclinadas a la parte derecha de la gráfica tienen un mayor Leverage.

Figura 29. Gráfico de Influencia para PC-5 de las 13 muestras orgánicas



En la gráfica de Leverage **Figura 30**, se puede evidenciar que no se presentan muestras con carácter negativo, ya que los valores de influencia son relativamente similares. No se observan muestras que sobrepasen el límite de influencia, por lo tanto no se tiene ninguna con alto Leverage que puedan perjudicar los futuros modelos y como se había dicho anteriormente también se logra ver con más claridad las muestras que tienen mayor Leverage (picos altamente pronunciados) como glicina, ácido adípico, antraceno y naftaleno.

Figura 30. Gráfico de Leverage para PC₅ en el modelo PCA



6.2 DESARROLLO DEL MODELO DE MÍNIMO CUADRADO PARCIALES (PLS)

Mediante el algoritmo de regresión PLS y el software THE UNSCRAMBLER versión 10.3 (CAMO) se evaluó el potencial de la espectroscopia MIR-ATR para predecir la relación H/C de muestras orgánicas.

Para el desarrollo de los modelos predictivo se tuvieron en cuenta dos etapas; la calibración y la validación. Sin embargo, debido al número de muestras con las

que se disponía para este análisis (13 muestras), se utilizó el total de ellas para construir el modelo de calibración y se usó la calibración cruzada para determinar la capacidad predictiva integral de cada uno de los modelos en espectroscopia raman y espectroscopia infrarroja.

La evaluación del modelo se realizó teniendo en cuenta los parámetros estadísticos RMSEC, RMSECV, RMSEP y la varianza explicada para encontrar el número óptimo de variables latentes.

A continuación se describe detalladamente el procedimiento que se empleó durante las etapas de calibración y validación de los modelos PLS de la propiedad de interés para las dos técnicas analíticas.

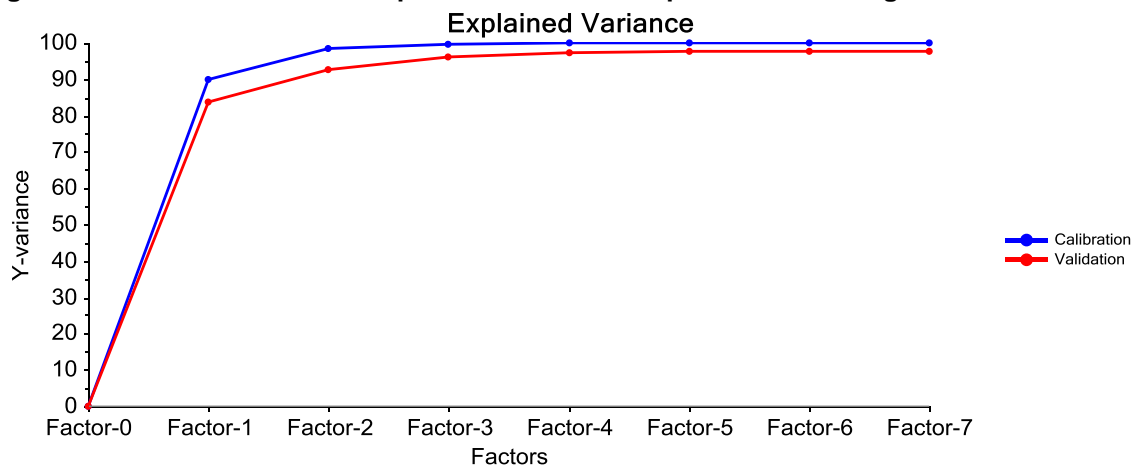
6.2.1 Calibración del modelo PLS para la espectroscopia Infrarroja (FTIR-ATR)

El modelo de predicción explica el 99,9% de la varianza de los datos tomando como máximo siete factores principales (**tabla 10**) (**figura 31**).

Tabla 10. Varianza explicada por los componentes principales de las muestras orgánicas para PLS

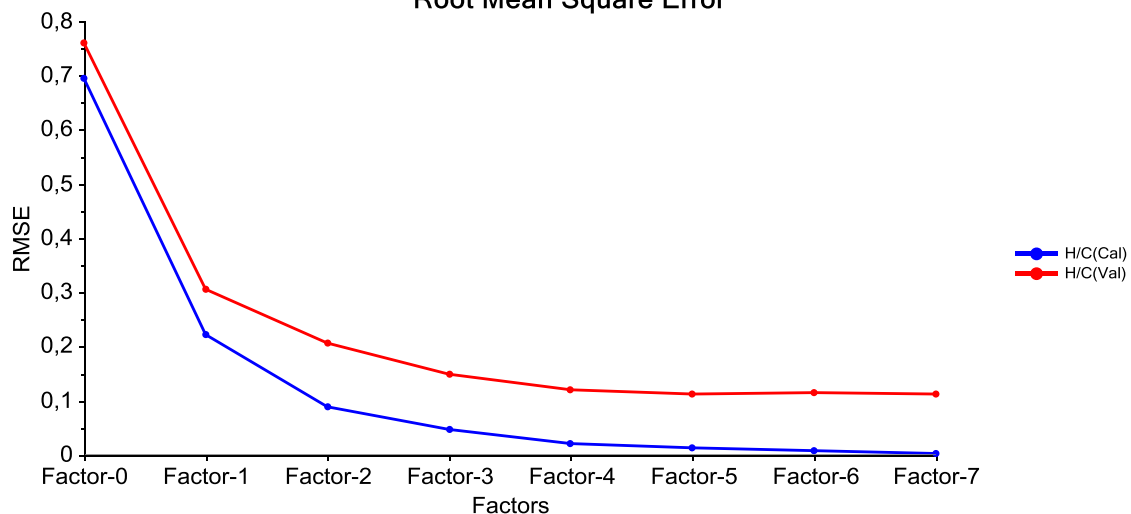
Componente Principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
<i>PC₁</i>	89,81	89,81
<i>PC₂</i>	8,51	98,32
<i>PC₃</i>	1,19	99,52
<i>PC₄</i>	0,39	99,91
<i>PC₅</i>	0,06	99,97
<i>PC₆</i>	0,02	99,99
<i>PC₇</i>	0,01	99,99

Figura 31. Gráfica de varianza explicada en el modelo para muestras orgánicas



En la **figura 32** se puede visualizar el número óptimo de componentes principales (PCs) para generar el modelo predictivo mediante la raíz cuadrada de los errores de calibración (RMSEC) y de validación cruzada (RMSEV) de color azul y rojo respectivamente. **RMSEC** es una medida de la desviación estándar de los residuales que se obtuvieron por la diferencia entre los valores de referencia y los predichos por el modelo para las muestras de calibración mientras que **RMSEV** se basa en un algoritmo iterativo que selecciona muestras dentro del grupo de calibración para desarrollar el modelo de predicción y luego evaluarlo sobre las muestras faltantes.

Figura 32. Error calculado en función del número de factores principales
Root Mean Square Error



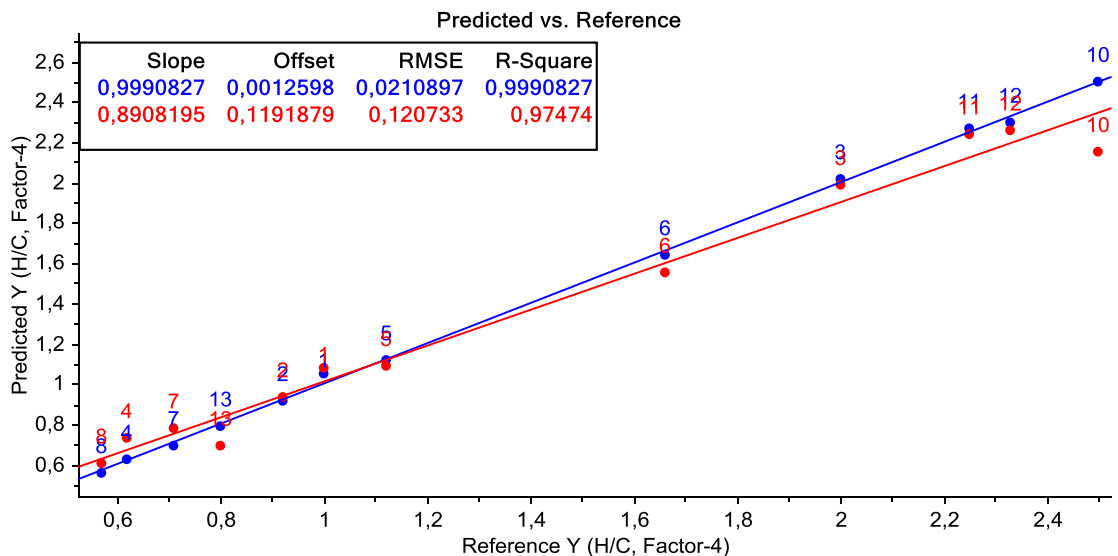
A medida que aumenta el factor, el error disminuye debido a que se hace menor la cantidad de variables analizadas, al retirar aquellas poco relevantes que no influyen sobre la propiedad de la muestra.

Los coeficientes de regresión son valores constantes (el peso) que se le da a cada una de las variables que se utilizaron en el modelo de predicción. Para el modelo PLS desarrollado con el espectro IR se encontró que las variables de mayor relevancia se encuentran en el rango de 2860 a 2849 cm^{-1} aproximadamente y dentro de este rango se encuentran las vibraciones simétricas y asimétricas de CH alifático y de 1404 a 1371 cm^{-1} manifiestan una absorción en la región de flexión C – H, esto es muy común en hidrocarburos de cadena ramificada como el Isooctano.

En el modelo PLS se eliminó la muestra número 9, permaneciendo estable el RMSEC y RMSEV ya que aunque se encontró adecuada para el modelo PLS, generaba error al modelo, ya que estaba un poco alejada de la recta de calibración generando un R^2 bajo y errores altos. El gráfico de predicción Vs referencia

(figura 32) es la confirmación de que el modelo es adecuado para predecir la relación H/C; en él encontramos dos líneas rectas que pasan a través de los puntos que representan las muestras orgánicas; la línea de validación representada de color rojo tiene un coeficiente de correlación de 0,97 con una error de 0,12 y la línea de color azul es la línea de calibración con una correlación de 0,99 y presenta un error RMSEC de 0,02 Se observa en la gráfica un punto discrepante que sobresale del resto de valores que representa a la muestra orgánica Glicina que aunque está situada en el extremo del intervalo en donde tiene más influencia en la afectación de la estimación de los coeficientes de regresión que si estuviera situada en el centro del intervalo; no se elimina este punto ya que al hacerlo se aumenta el error de predicción y además con pocas muestras para realizar el modelo retirar una de ellas no es conveniente.

Figura 33. Curvas de calibración y validación del modelo de predicción de la relación H/C



Los resultados hallados en esta etapa demostraron que el modelo para la predicción de la relación H/C en muestras orgánicas es satisfactorio para el rango de calibración de 0,57 a 2,5 ya que se obtuvieron bajos porcentajes de error (tabla 11).

Tabla 11. Validación del modelo de predicción

MUESTRAS	VALOR TEÓRICO	VALOR PREDICHO	RESIDUAL	% ERROR RELATIVO
<i>Ácido antranílico</i>	1,00	1,05	0,05	5,31
<i>Difenilamina</i>	0,92	0,91	0,007	0,76
<i>1,4 Dioxano</i>	2,00	2,02	0,02	0,85
<i>Isatina</i>	0,62	0,63	0,01	1,61
<i>Acetanilida</i>	1,12	1,12	0,003	0,27
<i>Ácido adipico</i>	1,66	1,63	0,02	12,65
<i>Antraceno</i>	0,71	0,69	0,01	1,97
<i>Antraquinona</i>	0,57	0,56	0,01	2,28
<i>Glicina</i>	2,50	2,50	0,001	0,04
<i>Isooctano</i>	2,25	2,27	0,02	0,8
<i>Hexano</i>	2,33	2,30	0,03	1,24
<i>Naftaleno</i>	0,80	0,79	0,01	1,37

El modelo se utilizó para la predicción con las mismas muestras usadas en la calibración (validación interna).

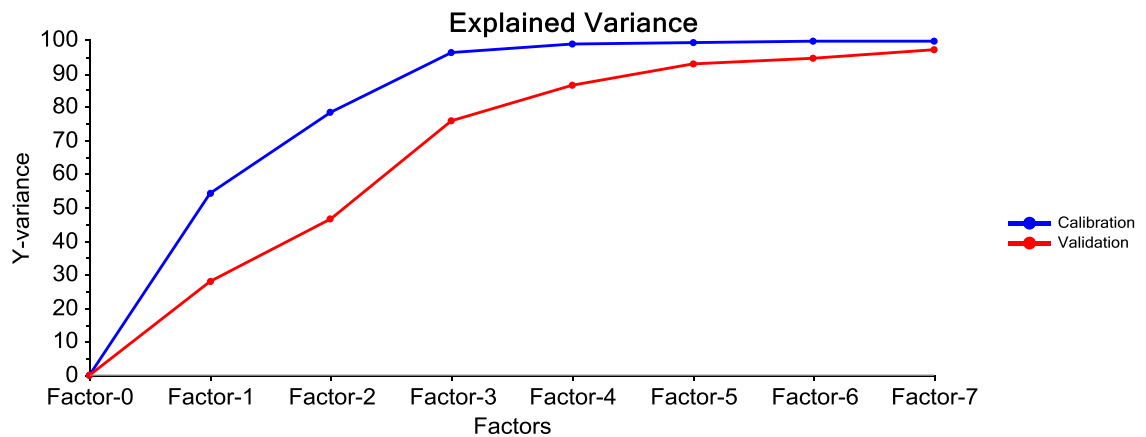
6.2.2 Calibración del modelo PLS para la Espectroscopia Raman. El modelo de predicción explica el 99,5% de la varianza de los datos tomando como máximo siete factores principales mostrados en la (tabla 12) (figura 34)

Tabla 12. Varianza explicada por los componentes principales de las muestras orgánicas

Componente Principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
<i>PC₁</i>	54,04	54,04
<i>PC₂</i>	24,46	78,50
<i>PC₃</i>	17,88	96,39

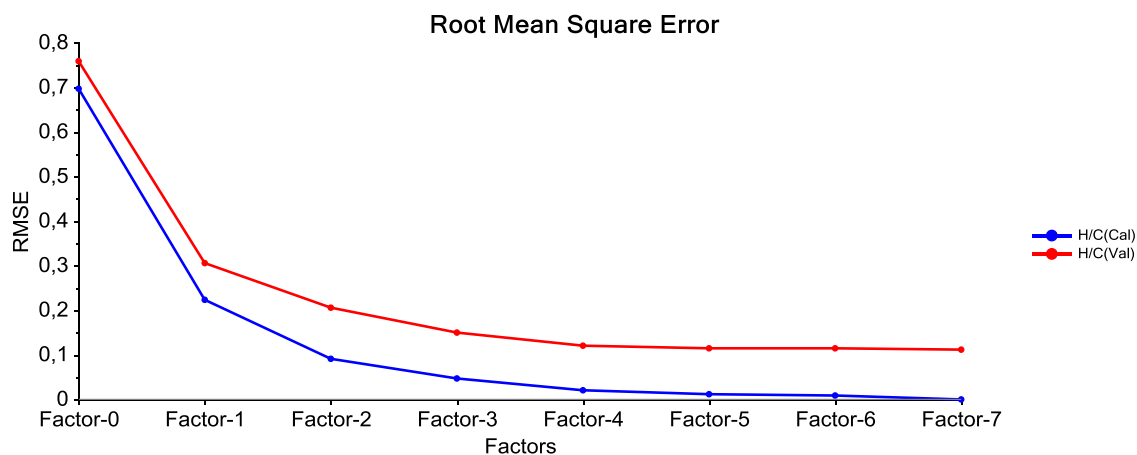
Componente Principal	Varianza Explicada (%)	Varianza Explicada Acumulada (%)
<i>PC₄</i>	2,53	98,92
<i>PC₅</i>	0,40	99,33
<i>PC₆</i>	0,13	99,45
<i>PC₇</i>	0,10	99,56

Figura 34. Gráfica de varianza explicada en el modelo para muestras orgánicas



La **figura 35** muestra el número óptimo de componentes principales necesarios para generar el modelo en función de los errores RMSEC y RMSEV, se observa que a medida que disminuye el error aumenta el número de factores principales. La línea de color rojo representa el error de validación y de color azul el error de calibración para cada componente principal.

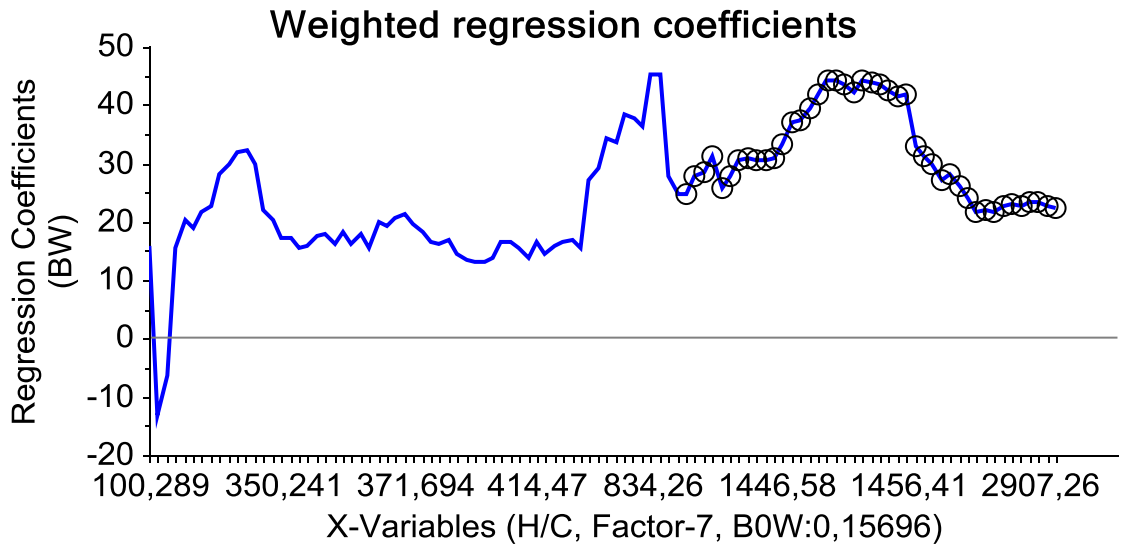
Figura 35. Error calculado en función del número de factores principales



En la **figura 36** se muestran los coeficientes de regresión para cada una de las variables que comprenden el rango entre 100,289 – 2907,26 cm^{-1} , Para el modelo PLS desarrollado con los espectros Raman. Se encontró que las variables de mayor relevancia (círculos negros) se encuentran en el rango de 2500 a 3000 cm^{-1} región que se asigna a bandas de grupos CH simétricos, asimétricos, alifáticos o aromáticos y en la región desde 3300 cm^{-1} a 1454 cm^{-1} aparece la señal característica de la vibración de enlaces nitrógeno-hidrógeno en los grupos amínicos (simétricos y asimétricos), aminas primarias y secundarias.

Las señales vibracionales de los enlaces CH_3 asimétricos y CH_2 se encontraron también en la región comprendida entre 1400 a 1470 cm^{-1}

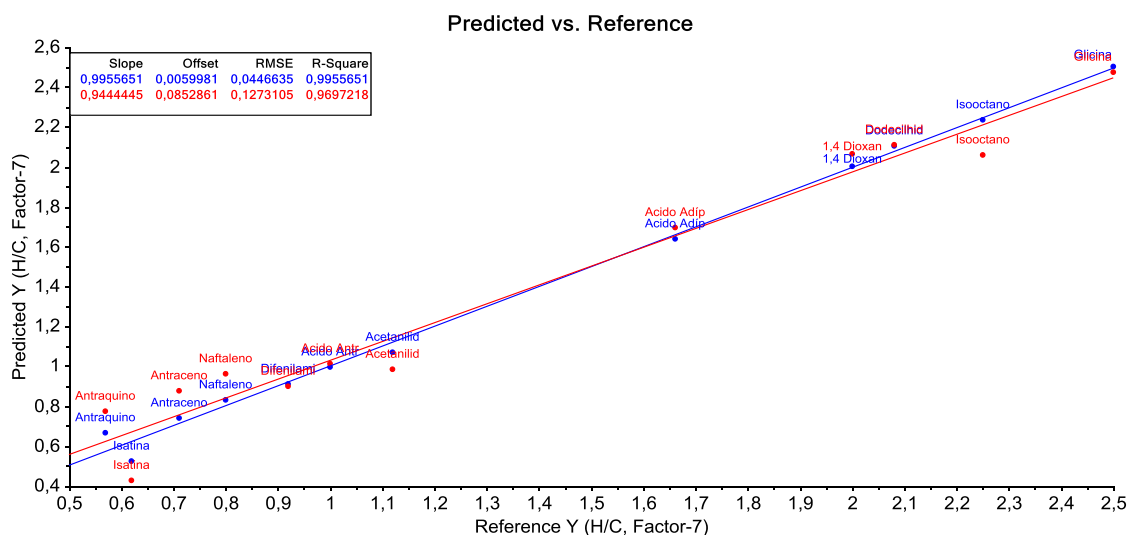
Figura 36. Coeficientes de regresión



La validación cruzada del modelo PLS para predecir la relación H/C en muestras orgánicas se realizó empleando el mismo grupo de muestras de calibración.

El coeficiente de correlación (R^2) entre los valores de referencia y de predicción hallado fue de 0,99 que representa la línea azul (calibración) y de 0,96 que hace referencia a la línea roja (validación) de la figura 37. El error estándar de predicción encontrado para el conjunto de muestras de validación (RMSEP) fue de 0,127.

Figura 37. Curvas de calibración y validación del modelo de predicción de la relación H/C



El modelo se desarrolló empleando 12 sustancias debido a la dificultad de encontrar muestras orgánicas de mayor o menor relación H/C de las empleadas en el rango de calibración de 0,57 a 2,5. Aún con este número reducido de muestras se puede decir que el modelo desarrollado para la técnica espectroscopia Raman es funcional en la determinación de la propiedad.

En la **tabla 13** se observa el porcentaje de error relativo que presentó el modelo usado para predecir la relación H/C por validación interna (muestras de calibración) encontrándose como valor máximo 14,9% correspondiente a la muestra Isatina.

Tabla 13. Validación del modelo de predicción

MUESTRAS	VALOR TEÓRICO	VALOR PREDICHO	RESIDUAL	% ERROR RELATIVO
<i>Ácido antranílico</i>	1,00	0,99	0,003	0,35
<i>Difenilamina</i>	0,92	0,90	0,01	1,17
<i>1,4 Dioxano</i>	2,00	2,00	0,004	0,21
<i>Isatina</i>	0,62	0,53	0,09	14,95
<i>Acetanilida</i>	1,12	1,07	0,05	4,35
<i>Ácido adipico</i>	1,66	1,64	0,02	1,31
<i>Antraceno</i>	0,71	0,74	0,03	4,27
<i>Antraquinona</i>	0,57	0,67	0,10	0,17
<i>Glicina</i>	2,50	2,50	0,02	0,08
<i>Isooctano</i>	2,25	2,24	0,01	0,61
<i>Naftaleno</i>	0,80	0,83	0,03	3,91

7. CONCLUSIONES

Se obtuvo dos modelos de predicción de la relación H/C de muestras orgánicas aplicable al rango de calibración de 0,57 a 2,5 con errores de calibración 0,02 para el modelo de IR y de 0,04 para el modelo de espectroscopia Raman.

Mediante el uso de técnicas espectroscópicas como FTIR-ATR y Raman en conjunto con métodos de calibración multivariable como PCA y PLS se puede predecir la relación H/C de manera económica y evitando pretratamiento a muestras orgánicas.

El análisis por componentes Principales permitió identificar agrupaciones naturales de los espectros de las muestras; en espectroscopia infrarroja se agruparon dependiendo de su estado físico mientras que espectroscopia Raman su agrupación se originó debido a su estructura molecular y el enlace del anillo aromático a átomos de carbono o de nitrógeno respectivamente.

Se necesitó emplear siete componentes principales para la elaboración del modelo predictivo en ambas técnicas debido a que las muestras son de diferente naturaleza química y por ende sus espectros difieren mucho el uno del otro. Siete es el número óptimo a incluir en el modelo de predicción por contener la mayor información de las variables originales. En Espectroscopia Infrarroja se explicó el 99,9 y en Espectroscopia Raman el 99,5.

Se determinó que las regiones del espectro de mayor importancia se encuentran para Espectroscopia Infrarroja en el rango de 2860 a 2849 cm^{-1} y 1404 a 1371 cm^{-1} y para Espectroscopia Raman en el rango de 2500 a 3000 cm^{-1} y 1454 cm^{-1} a 3300 cm^{-1}

Para la determinación de la relación H/C en muestras orgánicas se requiere analizar sustancias en las mismas condiciones en que se llevaron a cabo los análisis y con los parámetros ya establecidos en este trabajo investigativo para poder obtener una buena predicción tanto en Espectroscopia Infrarroja como en espectroscopia Raman.

8. RECOMENDACIONES

Se recomienda ampliar el rango de calibración usado en muestras orgánicas, robusteciendo los modelos desarrollados y evaluar continuamente los resultados manteniendo la información actualizada para todo tipo de sustancia de esta naturaleza química.

Es recomendable hacer la predicción con muestras externas a las usadas en la calibración para validar el método en el rango de 0,5 a 2,5.

Desarrollar modelos de predicción PLS para otras relaciones que incluyan los heteroátomos presentes en las moléculas, para poder determinar la fórmula empírica de compuestos orgánicos a través de éste método.

REFERENCIAS BIBLIOGRAFICAS

1. PhD, Luz Marina Jaramillo. *Química Orgánica General*. Santiago de Cali : Universidad del Valle, **2001**. Facultad de ciencias- Departamento de Química.
2. Sara Aldabe, Cecilia Bonazzola, Pedro Aramendía, Laura Lacreu. *Química 2, Química en acción*. Buenos Aires, Argentina : Colihue, **2004**. p. 246.
3. *The Chemical Formula Part II: Determination*. Doris, KOLB. 61635, Illinois : Central College, **1967**.
4. ERWIN BRAND, L. J. Saidel, W. H. Goldwater, Beatrice Kassel and F.J. Ryan. The Empirical Formula of B-Lactoglobulin. *CONTRIBUTION FROM THE DEPARTMENTS OF BIOCHEMISTRY AND ZOOLOGY, COLUMBIA UNIVERSITY*. Mayo 11 de **1945**, Vol. 67.
5. *Estimation of the Carbon / Hydrogen Ratio of Organic Compounds by Laser Pyrolysis*. Ristau, Nicholas E. Vanderborgh and William T. 8, Mexico : Department of chemistry, **1973**, Vol. 45. 1531.
6. *Empirical Formulae Studies of chlorofluorocarbons using gas chromatography coupled to atomic emission detection*. Nitin R Hadas, Peter C. Uden. Massachusetts : journal of Chromatography, **1999**.
7. Daniel T. Behe, Anders J. Lawaetz, Soren B. Engelsen, Marchen S Hviid, René Lametsch. *Accurate determination of endpoint temperature of cooked meat after storage by Raman spectroscopy and chemometrics*. Frederiksberg, Denmark : Elsevier, **2014**. Department of Food Science, Faculty of Science, University of Copenhagen.
8. Cyril Muehlethaler, Geneviene Massonnet, Pierre Esseiva. *The application of chemometrics on infrared and Ramanspectra as a tool for the forensic analysis of paints*. Switzerland : Elsevier, **2010**. School of criminal Sciences, Faculty of law, University of Lausanne, Batochime.
9. FRANCIS A, Carey. *Química orgánica*. University of Virginia : Mc Graw Hill, **2006**. pág. 506., sexta edición.
10. Skoog, D. A., Holler, J. H., Nieman, T. A. *Principios de Análisis Instrumental*. Madrid, España : McGraw Hill, **2001**. pág. 410. 5a Edición.
11. LESLY VIVIANA MELÉNDEZ CORREA., ADRIANA LACHE GARCÍA. *PREDICCIÓN DEL ANÁLISIS SARA DE CRUDOS COLOMBIANOS*. BUCARAMANGA : UNIVERSIDAD INDUSTRIAL DE SANTANDER, **2010**.

12. Rubinson, K. A., Rubinson, J. F. *Análisis Instrumental*. Madrid, España : Pearson Educación, S.A., **2001**. 5a Edición.
13. ISABEL Sierra., Alonso Damién Pérez Quintanilla., et al. *Análisis Instrumental*. España : Gesbiblo, S, L., **2010**. pág 70.
14. *Estudio de la Estructura de carbones Colombianos por FTIR*. Ospino, Enrique Mejía. , Colombia : Editorial académica española, **2012**.
15. APARICIO M, S. . *Metodologías analíticas basadas en espectroscopía de infrarrojo y calibración multivariante, aplicación a la industria petroquímica*. Tarragona. Tesis Doctoral : Universidad de Rovira y Virgili, **2002**. Departamento de Química analítica y química orgánica..
16. Barrera, Diego Fernando Rivera. *Reducción de la acidez del crudo tecanare con una solución alcoholica de 2-Metilimidazol y su cuantificación aplicando Espectroscopía FTIR-ATR y Métodos Quimiométricos*. Bucaramanga : Universidad Industrial de Santander, **2014**. Trabajo de grado.
17. *Infrared Spectroscopy*. Conley, Robert T. Boston : Allyn and Bacon, INC., **1972**, Vol. second edition.
18. *Infrared Spectroscopy for food quality analysis and control*. [aut. libro] DA-WEN Sun. USA : Elsevier, **2009**.
19. Pasto, Daniel J. y Johnson, Carl R. *Determinación de estructuras orgánicas*. Sevilla : Reverté, **2003**. pág. 168.
20. *Fundamentos de Espectroscopía Raman*. s.l. : Journal , **1930**.
21. *Aplicación de Técnicas Quimiométricas para el Desarrollo de nuevos Métodos Cinético-Espectrofotométricos de Análisis*. M., PORCEL. p. 11-15, Barcelona : Universidad Autónoma de Barcelona , **2001**.
22. *Multivariate calibration* . MARTENS, H y MARTENS, M. New York : Jhon Wiley & Sons., **1989**, Vol. P 438.
23. *Wavelength selection for simultaneous spectroscopic analysis, experimental and theoretical study*. XU L, Schechter. pag. 68, s.l. : Analytical Chemistry, **1996**.
24. *Multivariate calibration. What is in chemometrics for the analytical chemistry* *Chimica Acta*. R, BRO. issues 1-2. p. 185-194, Frederiksberg, Denmark : s.n., **2003**, Vol. 500.
25. *Some recent trends in the calibration literature. Chemometrics and intelligent Laboratory Systems*. GELADI, P. **2002**, Vols. 60, p. 211-224.

26. MARTENS, H y al., et. Multivariate analysis of quality and introduction. segunda edición. Editorial John Wiley & sons, **2001**, p.445.
27. WOLD, S. *Technometrics*. **1978**. Vol 20 P 397.
28. STONE, M J. *Staist Soc.* **1973**. Vol 36. P 111.
29. Stefany, RODRIGUEZ CamachoSidney. Predicción de la composición elemental de fracciones del proceso de DELAYED COKING de fondos de vacío de crudos Colombianos empleando Espectroscopía de fluorescencia y métodos Quimiométricos. Universidad Industrial de Santander, **2014**, Vol. Tesis de pregrado.
30. K.R., BOWALSKI Beebe. *introduction to multivariate calibration and analysis*. s.l. : Analytical Chemistry , **1987**. pág 59.
31. Belén, OTERO Reija. *Estudio estructural y dinámico de sistemas organizados mediante sondas fluorescentes* . s.l. : Universidad de Santiago de compostela, facultad de ciencias. Tesis de doctorado.
32. MILLER, J. N y MILLER, J. C. *Estadística y Quimiometría para química analítica*. España : Prentice Hall, **2002**. cuarta edición. P.595.
33. LOHMOLLER, J. B. *Das Programmsystem LVPLS für Pfadmodelle mit Latenten Variablen*. **1984**. ZA- Information Nr.15..
34. WOLD H. *Soft Modelling by Variables; the Non-Linear Iterative Partial Least Squares Approach, en Perspectives in Probability and Statistics, Paper in Honour of M. S. London* : Barlett, ed J. Gani, Academic Press, **1975**.
35. *The Unscramble 6.0 User's Guide*, CAMO AS, . Trondheim, Norway : s.n., **1996**.
36. GELADI P, KOWALSKI BR. *Partial least-squares regression: a tutorial*. s.l. : Analytica Chimica , **1986**. Acta, 185..
37. EGAN, J, W. y L.S., MORGAN. *Aanl. Chem.* **1998**. Vol 70. P 2372.
38. EISENHART, y otros. *Techniques of statistical analysis. H. Hotelling. Multivariate Quality Control*. New York : Mc Graw-Hil, **1947**.
40. HHORIBA, Ltd. <http://www.horiba.com/scientific/products/raman-spectroscopy/raman-spectrometers/raman-microscopes/details/labram-hr-evolution-17309/#>. [En línea] **1996-2015**. [Citado el: 10 de Febrero de **2015**.]

BIBLIOGRAFIA

ALDABE, Sara, BONAZZOLA, Cecilia. ARAMENDIA, Pedro. Laura Lacreu. Química 2, Química en acción. Buenos Aires, Argentina : Colihue, 2004. p. 246.

ALLYN AND BACON. Infrared Spectroscopy. Conley, Robert T. Boston :, INC., 1972, Vol. second edition.

APARICIO M, S. . Metodologías analíticas basadas en espectroscopía de infrarrojo y calibración multivariante, aplicación a la industria petroquímica. Tarragona. Tesis Doctoral : Universidad de Rovira y Virgili, 2002. Departamento de Química analítica y química orgánica..

BEHE, Daniel T. LAWAETZ, Anders J.ENGELSEN, Soren B., MARCHEN S HVIID, LAMETSCH, René. Accurate determination of endpoint temperature of cooked meat after storage by Raman spectroscopy and chemometrics. Frederiksbergc, Denmark : Elsevier, 2014. Department of Food Science, Faculty of Science, University of Copenhagen.

BOWALSKI Beebe, K.R.,. introduction to multivariate calibration and analysis. s.l. : Analytical Chemistry , 1987. pág 59.

BRAND, Erwin, L. J. SAIDEL, W. H. GOLDWATER, KASSEL Beatrice, and F.J. RYAN. The Empirical Formula of B-Lactoglobulin. Contribution from the departments of Biochemistry and Zoology, Columbia University. Mayo 11 de 1945, Vol. 67.

DA-WEN Sun. Infrared Spectroscopy for food quality analysis and control. [aut. libro]. USA : Elsevier, 2009.

DENMARK, Frederiksberg. Multivariate calibration. What is in chemometrics for the analytical chemistry analytica Chimica Acta. R, BRO. issues 1-2. p. 185-194: s.n., 2003, Vol. 500.

EGAN, J, W. y L.S., MORGAN. Anal. Chem. 1998. Vol 70. P 2372.

EISENHART, y otros. Techniques of statistical analysis. H. Hotelling. Multivariate Quality Control. New York : Mc Graw-Hil, 1947.

FRANCIS A, Carey. Química orgánica. University of Virginia : Mc Graw Hill, 2006. pág. 506., sexta edición.

GELADI P, KOWALSKI BR. Partial least-squares regression: a tutorial. s.l. : Analytica Chimica , 1986. Acta, 185..

HADAS, Nitin R., UDEN, Peter C. Empirical Formulae Studies of chlorofluorocarbons using gas chromatography coupled to atomic emission detection. Massachusetts : journal of Chromatography, 1999.

HHORIBA, Ltd. <http://www.horiba.com/scientific/products/raman-spectroscopy/raman-spectrometers/raman-microscopes/details/labram-hr-evolution-17309/#>. [En línea] 1996-2015. [Citado el: 10 de Febrero de 2015.]

JARAMILLO, Luz Marina. PhD. Química Orgánica General. Santiago de Cali : Universidad del Valle, 2001. Facultad de ciencias- Departamento de Química.

KOLB, Doris. The Chemical Formula Part II: Determination.. 61635, Illinois : Central College, 1967.

LOHMOLLER, J. B. Das Programmsystem LVPLS für Pfadmodelle mit Latenten Variablen. 1984. ZA- Information Nr.15..

MARTENS, H y MARTENS, M Multivariate calibration .. New York : Jhon Wiley & Sons., 1989, Vol. P 438.

MARTENS, H y al., et. Multivariate analysis of quality and introduction. segunda edición. Editorial John Wiley & sons, 2001, p.445.

MEJIA OSPINO, Enrique. Estudio de la Estructura de carbones Colombianos por FTIR.. , Colombia : Editorial académica española, 2012.

MELÉNDEZ CORREA, Lesly Viviana, LACHE GARCIA, Adriana. Predicción del análisis sara de crudos colombianos. Bucaramanga : Universidad Industrial de Santander, 2010.

MILLER, J. N y MILLER, J. C. Estadística y Quimiometría para química analítica. España : Prentice Hall, 2002. cuarta edición. P.595.

MUEHLETHALER, Cyril, MASSONNET, Geneviene, ESSEIVA Pierre. The application of chemometrics on infrared and Ramanspectra as a tool for the forensic analysis of paints. Switzerland : Elsevier, 2010. School of criminal Sciences, Faculty of law, University of Lausanne, Batochime.

OTERO REIJA, Belén. Estudio estructural y dinámico de sistemas organizados mediante sondas fluorescentes . s.l. : Universidad de Santiago de compostela, facultad de ciencias. Tesis de doctorado.

PASTO, Daniel J. y JOHNSON, Carl R. Determinación de estructuras orgánicas. Sevilla : Reverté, 2003. pág. 168.

PORCEL. Aplicación de Técnicas Quimiométricas para el Desarrollo de nuevos Métodos Cinético-Espectrofotométricos de Análisis. M., p. 11-15, Barcelona : Universidad Autónoma de Barcelona , 2001.

RISTAU, Nichoolas. E. VANDERBORGH and William. Estimation of the Carbon / Hydrogen Ratio of Organic Compounds by Laser Pyrolysis. 8, Mexico : Department of chemistry, 1973, Vol. 45. 1531.

RIVERA BARRERA, Diego Fernando. Reducción de la acidez del crudo tecanare con una solución alcoholica de 2-Metilimidazol y su cuantificación aplicando Espectroscopía FTIR-

ATR y Métodos Quimiométricos. Bucaramanga : Universidad Industrial de Santander, 2014. Trabajo de grado.

RODRIGUEZ, Stefany. CamachoSidney. Predicción de la composición elemental de fracciones del proceso de DELAYED COKING de fondos de vacío de crudos Colombianos empleando Espectroscopía de fluorescencia y métodos Quimiométricos. Universidad Industrial de Santander, 2014, Vol. Tesis de pregrado.

RUBINSON, K. A., RUBINSON, J. F. Análisis Instrumental. Madrid, España : Pearson Educación, S.A., 2001. 5a Edición.

SCHECHTER, Xu L. Wavelength selection for simultaneous spectroscopic analysis, experimental and theoretical study.. pag. 68, s.l. : Analytical Chemistry, 1996.

SIERRA, Isabel, PÉREZ QUINTANILLA, Alonso Damién., et al. Análisis Instrumental. España : Gesbiblo, S, L., 2010. pág 70.

Some recent trends in the calibration literature. Chemometrics and intelligent Laboratory Systems. GELADI, P. 2002, Vols. 60, p. 211-224.

SKOOG, D. A., Holler, J. H., NIEMAN, T. A. Principios de Análisis Instrumental. Madrid, España : McGraw Hill, 2001. pág. 410. 5a Edición.

The Unscramble 6.0 User's Guide, CAMO AS, . Trondheim, Norway : s.n., 1996.

WOLD H. Soft Modelling by Variables; the Non-Linear Iterative Partial Least Squares Approach, en Perspectives in Probability and Statistics, Paper in Honour of M. S. London : Barlett, ed J. Gani, Academic Press, 1975.