Universidad del Atlántico	**CÓDIGO:** FOR-DO-109
	VERSIÓN: 0
	FECHA: 03/06/2020

AUTORIZACIÓN DE LOS AUTORES PARA LA CONSULTA, LA REPRODUCCIÓN PARCIAL O TOTAL, Y PUBLICACIÓN ELECTRÓNICA DEL TEXTO COMPLETO

Autor1

Puerto Colombia, **22 de Abril de 2020**

Señores
DEPARTAMENTO DE BIBLIOTECAS
Universidad del Atlántico

Asunto: Autorización Trabajo de Grado

Cordial saludo,

Yo, **MÓNICA TATIANA RUEDA SÁNCHEZ**, identificado(a) con **C.C. No. 1.140.871.553** de **BARRANQUILLA**, autor(a) del trabajo de grado titulado **DESARROLLO DE UN MODELO DE CALIBRACIÓN MULTIVARIADO PARA LA DETERMINACIÓN Y CUANTIFICACIÓN DE ETANOL Y METANOL EN GASOLINA UTILIZANDO LA TÉCNICA DE INFRARROJO CERCANO ACOPLADO A ALGORITMO GENÉTICO** presentado y aprobado en el año **2020** como requisito para optar al título Profesional de **QUIMICO**; autorizo al Departamento de Bibliotecas de la Universidad del Atlántico para que, con fines académicos, la producción académica, literaria, intelectual de la Universidad del Atlántico sea divulgada a nivel nacional e internacional a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios del Departamento de Bibliotecas de la Universidad del Atlántico pueden consultar el contenido de este trabajo de grado en la página Web institucional, en el Repositorio Digital y en las redes de información del país y del exterior, con las cuales tenga convenio la Universidad del Atlántico.
- Permitir consulta, reproducción y citación a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato CD-ROM o digital desde Internet, Intranet, etc., y en general para cualquier formato conocido o por conocer.

Esto de conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “Los derechos morales sobre el trabajo son propiedad de los autores”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

Atentamente,

Firma Monica Tatiana Rueda Sanchez
MÓNICA TATIANA RUEDA SÁNCHEZ
C.C. No. 1.140.871.553 de BARRANQUILLA

Universidad del Atlántico	**CÓDIGO:** FOR-DO-109
	VERSIÓN: 0
	FECHA: 03/06/2020

AUTORIZACIÓN DE LOS AUTORES PARA LA CONSULTA, LA REPRODUCCIÓN PARCIAL O TOTAL, Y PUBLICACIÓN ELECTRÓNICA DEL TEXTO COMPLETO

Autor2

Puerto Colombia, **22 de Abril de 2020**

Señores
DEPARTAMENTO DE BIBLIOTECAS
Universidad del Atlántico

Asunto: Autorización Trabajo de Grado

Cordial saludo,

Yo, **ARLINE JOILL ROMERO ARROYO**, identificado(a) con **C.C. No. 1.140.882.849** de **BARRANQUILLA**, autor(a) del trabajo de grado titulado **DESARROLLO DE UN MODELO DE CALIBRACIÓN MULTIVARIADO PARA LA DETERMINACIÓN Y CUANTIFICACIÓN DE ETANOL Y METANOL EN GASOLINA UTILIZANDO LA TÉCNICA DE INFRARROJO CERCANO ACOPLADO A ALGORITMO GENÉTICO** presentado y aprobado en el año **2020** como requisito para optar al título Profesional de **QUIMICO**; autorizo al Departamento de Bibliotecas de la Universidad del Atlántico para que, con fines académicos, la producción académica, literaria, intelectual de la Universidad del Atlántico sea divulgada a nivel nacional e internacional a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios del Departamento de Bibliotecas de la Universidad del Atlántico pueden consultar el contenido de este trabajo de grado en la página Web institucional, en el Repositorio Digital y en las redes de información del país y del exterior, con las cuales tenga convenio la Universidad del Atlántico.
- Permitir consulta, reproducción y citación a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato CD-ROM o digital desde Internet, Intranet, etc., y en general para cualquier formato conocido o por conocer.

Esto de conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “Los derechos morales sobre el trabajo son propiedad de los autores”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

Atentamente,

Firma
ARLINE JOILL ROMERO ARROYO
C.C. No. 1.140.882.849 de BARRANQUILLA

Universidad del Atlántico	CÓDIGO: FOR-DO-110
	VERSIÓN: 01
	FECHA: 02/DIC/2020

DECLARACIÓN DE AUSENCIA DE PLAGIO EN TRABAJO ACADÉMICO PARA GRADO

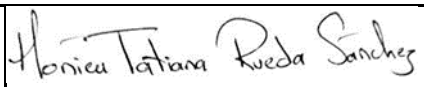
Este documento debe ser diligenciado de manera clara y completa, sin tachaduras o enmendaduras y las firmas consignadas deben corresponder al (los) autor (es) identificado en el mismo.

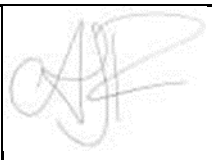
Puerto Colombia, **22 de Abril de 2020**

Una vez obtenido el visto bueno del director del trabajo y los evaluadores, presento al **Departamento de Bibliotecas** el resultado académico de mi formación profesional o posgradual. Asimismo, declaro y entiendo lo siguiente:

- El trabajo académico es original y se realizó sin violar o usurpar derechos de autor de terceros, en consecuencia, la obra es de mi exclusiva autoría y detento la titularidad sobre la misma.
- Asumo total responsabilidad por el contenido del trabajo académico.
- Eximo a la Universidad del Atlántico, quien actúa como un tercero de buena fe, contra cualquier daño o perjuicio originado en la reclamación de los derechos de este documento, por parte de terceros.
- Las fuentes citadas han sido debidamente referenciadas en el mismo.
- El (los) autor (es) declara (n) que conoce (n) lo consignado en el trabajo académico debido a que contribuyeron en su elaboración y aprobaron esta versión adjunta.

Título del trabajo académico:	**DESARROLLO DE UN MODELO DE CALIBRACIÓN MULTIVARIADO PARA LA DETERMINACIÓN Y CUANTIFICACIÓN DE ETANOL Y METANOL EN GASOLINA UTILIZANDO LA TÉCNICA DE INFRARROJO CERCANO ACOPLADO A ALGORITMO GENÉTICO**
Programa académico:	**QUÍMICA**

Firma de Autor 1:								
Nombres y Apellidos:	**MÓNICA TATIANA RUEDA SÁNCHEZ**							
Documento de Identificación:	**CC**	**X**	**CE**		**PA**		**Número:**	**1.140.871.553**
Nacionalidad:					**Lugar de residencia:**			
Dirección de residencia:								
Teléfono:					**Celular:**			

Firma de Autor 2:								
Nombres y Apellidos:	**ARLINE JOILL ROMERO ARROYO**							
Documento de Identificación:	**CC**	**X**	**CE**		**PA**		**Número:**	**1.140.882.849**
Nacionalidad:					**Lugar de residencia:**			
Dirección de residencia:								
Teléfono:					**Celular:**			

Universidad del Atlántico	CÓDIGO: FOR-DO-110
	VERSIÓN: 01
	FECHA: 02/DIC/2020

DECLARACIÓN DE AUSENCIA DE PLAGIO EN TRABAJO ACADÉMICO PARA GRADO

TÍTULO COMPLETO DEL TRABAJO DE GRADO	**DESARROLLO DE UN MODELO DE CALIBRACIÓN MULTIVARIADO PARA LA DETERMINACIÓN Y CUANTIFICACIÓN DE ETANOL Y METANOL EN GASOLINA UTILIZANDO LA TÉCNICA DE INFRARROJO CERCANO ACOPLADO A ALGORITMO GENÉTICO**
AUTOR(A) (ES)	**MÓNICA TATIANA RUEDA SÁNCHEZ ARLINE JOILL ROMERO ARROYO**
DIRECTOR (A)	**JORGE ROPERO VEGA**
CO-DIRECTOR (A)	**MÓNICA ISABEL GARCÍA AGUDELO**
JURADOS	**MARIO ALBERTO ROMERO CALONGE CARLOS ALBERTO TOLOZA TOLOZA**
TRABAJO DE GRADO PARA OPTAR AL TITULO DE	**QUÍMICO**
PROGRAMA	**QUÍMICA**
PREGRADO / POSTGRADO	**PREGRADO**
FACULTAD	**CIENCIAS BÁSICAS**
SEDE INSTITUCIONAL	**PUERTO COLOMBIA**
AÑO DE PRESENTACIÓN DEL TRABAJO DE GRADO	**2020**
NÚMERO DE PÁGINAS	**74**
TIPO DE ILUSTRACIONES	**NO APLICA**
MATERIAL ANEXO (VÍDEO, AUDIO, MULTIMEDIA O PRODUCCIÓN ELECTRÓNICA)	**NO APLICA**
PREMIO O RECONOMIENTO	**NO APLICA**

DESARROLLO DE UN MODELO DE CALIBRACIÓN MULTIVARIADO PARA LA DETERMINACIÓN Y CUANTIFICACIÓN DE ETANOL Y METANOL EN GASOLINA UTILIZANDO LA TÉCNICA DE INFRARROJO CERCANO ACOPLADO A ALGORITMO GENÉTICO

ARLINE JOILL ROMERO ARROYO
MÓNICA TATIANA RUEDA SÁNCHEZ

UNIVERSIDAD DEL ATLÁNTICO
FACULTAD DE CIENCIAS BÁSICAS
PROGRAMA DE QUÍMICA
BARRANQUILLA – ATLÁNTICO
2019

DESARROLLO DE UN MODELO DE CALIBRACIÓN MULTIVARIADO PARA LA DETERMINACIÓN Y CUANTIFICACIÓN DE ETANOL Y METANOL EN GASOLINA UTILIZANDO LA TÉCNICA DE INFRARROJO CERCANO ACOPLADO A ALGORITMO GENÉTICO

ARLINE JOILL ROMERO ARROYO
MÓNICA TATIANA RUEDA SÁNCHEZ

Trabajo de tesis para optar al título de QUÍMICO

DIRECTOR
JORGE ROPERO VEGA, Ph.D
Químico
Doctorado en Química de Materiales

CO-DIRECTOR
MÓNICA ISABEL GARCÍA AGUDELO
Químico

UNIVERSIDAD DEL ATLÁNTICO
FACULTAD DE CIENCIAS BÁSICAS
PROGRAMA DE QUÍMICA
BARRANQUILLA – ATLÁNTICO

NOTA DE ACEPTACIÓN: ____

OBSERVACIONES:

_____

_____

_____

_____

_____

_____

_____	_____
Ph.D. Jorge Ropero Vega	**Mónica García Agudelo**
Director	**Co-Director**

_____

MSc. Dency José Pachecho López
Coordinado Comité Trabajo de Grado

_____	_____
Ph.D. Mario Romero Calonge	**Dr. Carlos Toloza Toloza**
Jurado	**Jurado**

Barranquilla, 7 de noviembre 2019

AGRADECIMIENTOS

Deseo manifestar mis agradecimientos a:

A Dios por habernos dado la sabiduría y la fortaleza para llegar al punto donde hoy nos encontramos, además por impulsarnos cada día a superarnos.

A nuestro director de proyecto el Dr. **Jorge Ropero** por habernos dado la oportunidad de realizar este trabajo y abrirnos las puertas del magnífico mundo de la espectroscopía NIR y la Quimiometría.

A nuestra Co-directora **Mónica García** por su tiempo y compartir sus conocimientos con nosotras para lograr un mejor trabajo.

A **Jorge Hernández** por proveernos información necesaria para el desarrollo de nuestro trabajo.

A **Procaps** S.A. Barranquilla, Colombia, por la capacitación teórico-práctica y la ayuda brindada por adquisición de los espectros de infrarrojo cercano.

A **Dahgna Páez** y **Alejandro Romero** por su ayuda con los softwares quimiométricos y estadísticos utilizados.

Y en especial a nuestras familias por apoyarnos en cada paso de este proceso para culminar nuestro Tesis de Pregrado.

LISTA DE FIGURAS

Pág.

LISTA DE TABLA

Pág.

CONTENIDO

RESUMEN

En este trabajo se construyó un modelo de calibración multivariado el cual permite determinar y cuantificar el porcentaje de etanol y metanol en la gasolina, utilizando como método de selección espectral el algoritmo genético quien indica las longitudes de onda donde se encuentra la mayor variación de las muestras. El procedimiento consistió en preparar mezclas con diferentes proporciones (0-15%) de metanol y etanol siendo lo restante gasolina, cumpliendo con los porcentajes reales establecidos en la Resolución 40185 del 2018 emitida por el ministerio de Minas y Energías. Las muestras fueron analizadas por espectroscopía infrarrojo cercano (NIR). A la serie de datos obtenidos se le aplicó algoritmo genético en el software estadístico R Project versión 3.6.0 obteniendo así la zona espectral con la cual se desarrolló el modelo. Por último, se hizo la Regresión de Componente Principal (PCA) el cual presentó relaciones entre las muestras y una variabilidad del 95% para el primer componente. También se hizo uso de la Regresión por Mínimos Cuadrado Parciales (PLS) quien mostró a través del Error Medio Cuadrático (RMSE) (0,423) y el R^2 (0,999) para etanol y, para metanol el RMSE (0,245) y el R^2 (0,999) que el mejor modelo de predicción fue cuando se le aplicó corrección de línea base, algoritmo genético y con cuatro (4) factores.

Palabras claves: algoritmo genético, regresión de mínimos cuadrados (PLS) y regresión de componente principal (PCA)

ABSTRACT

In this work, a multivariate calibration model was constructed which allows to determine and quantify the percentage of ethanol and methanol in gasoline, using as a method of spectral selection the genetic algorithm who indicates the wavelengths where the greatest variation of the samples is found. The procedure consists in preparing mixtures with different proportions (0-15%) of methanol and ethanol being the remainder of gasoline, complying with the real percentages established in Resolution 40185 of 2018 issued by the Ministerio de minas y Energía. The samples were analyzed by near infrared spectroscopy (NIR). The genetic algorithm in the statistical software R Project version 3.6.0 is applied to the series of data obtained, thus obtaining the spectral zone with the quality of the model. Finally, the Principal Component Regression (PCA) was made, which had relations between the samples and a 95% variability for the first component. The Partial Least Square Regression (PLS) was also used, which showed through RMSE (0.423) and R2 (0.999) for ethanol and, for methanol, RMSE (0.245) and R2 (0.999) as the best model Prediction was when the baseline correction, genetic algorithm and four (4) factors were applied.

Keywords: genetic algorithm, partial least squares regression (PLS) and principal component regression (PCA).

1. INTRODUCCIÓN

La gasolina producida a principios del siglo XX contenía hidrocarburos de hasta dieciséis (16) átomos de carbono, lo que producía para el combustible un bajo número de octano. Durante el último siglo, se desarrolló un nuevo proceso de producción para obtener un aumento en el octanaje del combustible el cual podría ser usado en autos con motores de alta compresión. Sin embargo, este nuevo procedimiento implicaba el uso de aditivos como el tetraetilo de plomo y el metilterbutileter (MTBE) quienes contribuyen altamente a la contaminación atmosférica y al poco rendimiento de los motores [1].

El reciente interés por el bioetanol como aditivo para la gasolina es debido precisamente a la disminución de gases de emisión provenientes del sector transporte, así mismo, se reduce la dependencia en el uso de los combustibles fósiles. Además, éste mantiene el octanaje óptimo para el buen funcionamiento del motor [2]. Otro beneficio del bioetanol es su procedencia de fuentes renovables como, por ejemplo, su síntesis a partir de la caña de azúcar [3].

En Colombia, el Ministerio de Minas y Energías estableció desde el año 2001 el uso de alcoholes carburantes como el etanol en mezcla con la gasolina. Ésto se plasmó en la ley 693 de ese mismo año, con un porcentaje correspondiente al etanol del 5%, siendo el porcentaje restante perteneciente a la gasolina [4]. Actualmente, la resolución 40185 del 2018 declara que los porcentajes en la mezcla gasolina/etanol son 90%/10% respectivamente [5].

Tanto la producción como la distribución del etanol van estrechamente ligados a la posibilidad de adulteración de esta mezcla con alcoholes de propiedades químicas y físicas similares como el metanol, debido a que químicamente se diferencia del etanol solo por un carbono. Razón por la cual, los métodos de cuantificación para estos alcoholes en dicha mezcla son importantes para así lograr establecer la calidad de las mismas.

La espectroscopía de infrarrojo cercano o NIR es una técnica analítica no destructiva la cual es utilizada en la determinación de los productos resultantes de la refinería del petróleo debido a su alto contenido de hidrocarburos [6][7]. Ésta técnica en unión con la quimiometría resultan bastante útil para la detección y cuantificación de etanol y metanol en mezclas con gasolina aplicadas en este caso, a través del desarrollo de modelos de calibración multivariados para los cuales se hace uso del algoritmo genético quien indica las longitudes de onda donde se encuentra la mayor variabilidad de las muestras.

En el presente trabajo, se aplica la espectroscopía NIR junto con el algoritmo genético y la calibración multivariada (PLS), para la determinación simultánea de metanol y etanol en gasolina, verificando la robustez del modelo de calibración cuando se someta a la validación cruzada.

2. MARCO DE REFERENCIA Y ESTADO DEL ARTE

2.1 ANTECEDENTES

En los últimos años, algunos factores sociales, políticos, económicos, tecnológicos y ambientales han llevado al desarrollo de fuentes alternativas de energía a gran escala, entre los cuales se encuentran los biocombustibles, más específicamente el bioetanol y el biodiesel. De esta manera el bioetanol se convierte en una alternativa real para el petróleo y se proyecta como un gran generador del desarrollo del campo colombiano en el siglo XXI.

Con referente al tema de energía alternativa, en la mayoría de los casos se hace referencia a las llamadas fuentes convencionales para la obtención de tal recurso, entre los cuales se encuentra el petróleo, carbón y/o gas natural. Estas fuentes se caracterizan por ser recursos naturales no renovables [8].

Como alternativa para superar los problemas de orden económico, social, político y ambiental generados con el indiscriminado uso de los combustibles de origen fósil, actualmente las investigaciones apuntan al desarrollo de fuentes de energía renovables o también conocidos como biocombustibles. El uso de éstos permite a los países la diversificación de la canasta energética y hace menos dependiente a este sector con respecto al uso de combustibles fósiles. Los biocombustibles pueden tener efectos positivos sobre el medio ambiente al disminuir la emisión de gases de efecto invernadero, además de traer consigo impacto favorable sobre el desarrollo rural de los países.

El gobierno nacional, mediante la expedición de la ley 693 del 12 se septiembre de 2001, estableció que a partir del año 2005 la gasolina colombiana debería tener elementos oxigenados que disminuyeran las emisiones para el ambiente. En Colombia la producción industrial de bioetanol da inicio en el año 2005 y, se optó

por la utilización de caña de azúcar debido a que esta se encuentra consolidada en el país y presenta una mayor eficiencia energética frente a otras materias primas a partir de las cuales se produce etanol como los cereales, la remolacha, el maíz o el trigo [9, 10, 11].

Colombia es uno de los países con mayor producción de caña de azúcar y, esta privilegiada posición con respecto a otros países fue la razón por la cual se elige este cultivo como la materia prima que ofrece las mejores posibilidades para la producción de bioetanol en Colombia, esto se puede observar en la Figura 1 [12].

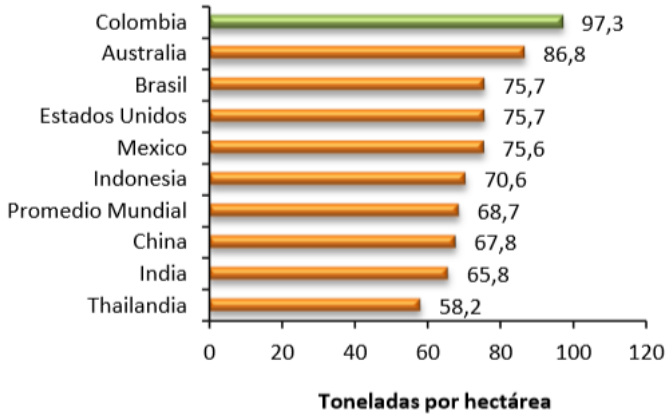


Figura 1. Gráfico sobre el rendimiento promedio en la producción de la caña de azúcar.

A nivel mundial la materia prima más empleada para la obtención de etanol son los cereales (especialmente el maíz) Además, son los más usados en Europa y norte américa.

En Brasil, la materia prima más utilizada para la producción de etanol es la caña de azúcar.

La Figura 2 muestra la producción de etanol por país entre el 2007 y el 2017. Estados unidos es el país con mayor producción de etanol a nivel mundial, habiendo producido casi 16 mil millones de galones solo en el 2017. Estados unidos junto con

Brasil producen el 85% de etanol del mundo, comparado con Colombia el cual produce el 0,36% del etanol a nivel mundial.

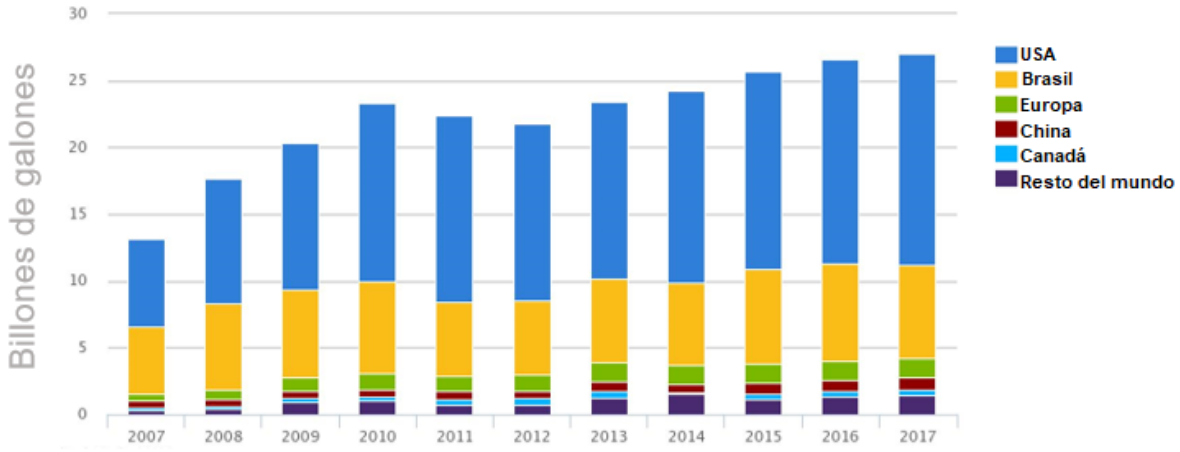


Figura 2.Gráfico sobre la producción de etanol a nivel mundial

2.1.1 Uso obligatorio de bioetanol en la gasolina en Colombia

La ley 693 del 19 de septiembre del 2001, expresa que la gasolina utilizada en el país, en especial, en los centros urbanos de más de 500.000 habitantes debe contener componentes oxigenados tales como alcoholes carburantes en la cantidad y calidad que establezca el Ministerio de Minas y Energía. Lo anterior se fundamenta de acuerdo a la reglamentación sobre el control de emisiones derivadas del uso combustibles fósiles y los requerimientos de saneamiento ambiental que dicte el Ministerio del Medio Ambiente para cada región del país [13].

El decreto 4892 de diciembre del 2011, en su artículo 1°, estipula que la gasolina de motor debe tener porcentajes de mezcla obligatoria que varíen entre el 8% y el 10% de la mezcla de alcohol carburante en base volumétrica (E-8 E-10 corriente y extra) [14].

2.1.2 Mezclas de bioetanol/gasolina en Colombia

El uso del bioetanol en la matriz energética colombiana empezó con la ley nacional 693 del 2001, la cual establece que la gasolina debe tener componentes oxigenados

tales como alcoholes carburantes, en la cantidad y calidad que establezca el ministerio de Minas y Energía gasolina motor con porcentajes de mezcla obligatoria que varían entre el 8% y el 10% de mezcla de alcohol carburante en base volumétrica (E-8- E-10 corriente y extra) [3].

Mediante la Resolución 40527 del 7 de junio de 2017 se restablecieron los porcentajes de mezcla de gasolina para motor con alcohol carburante de la siguiente manera: i) a partir del 9 de junio de 2017: seis por ciento (6%) de alcohol carburante con noventa y cuatro por ciento (94%) de gasolina motor, denominada E-6, para los departamentos ubicados en la zona sur, suroccidente, centro, oriente y nororiente del país: ii) a partir del 10 de julio 2017: ocho por ciento (8%) de alcohol carburante con un noventa y dos por ciento (92%) de gasolina motor, denominada E-8, para los departamentos ubicados en la zona sur, suroccidente, centro, oriente, norte y nororiente del país [3].

De acuerdo con el concepto técnico emitido por la Dirección de Hidrocarburos del Ministerio de Minas y Energía, con radicado Nº 2018014350 del 27 de febrero de 2018, se definió con base en el estimado de producción y la proyección del año 2018, demostrando la viabilidad y capacidad de sostener una mezcla del 10% de alcohol carburante para todas aquellas zonas a nivel nacional que cuentan una mezcla del 8% [3].

Actualmente según el Ministerio de Minas y Energía, la composición de las mezclas bioetanol/gasolina en el territorio nacional se encuentra en porcentaje del 10% para etanol al cual se le denomina E10. Esta norma rige la mayoría del territorio nacional como se muestra en la Figura 3.

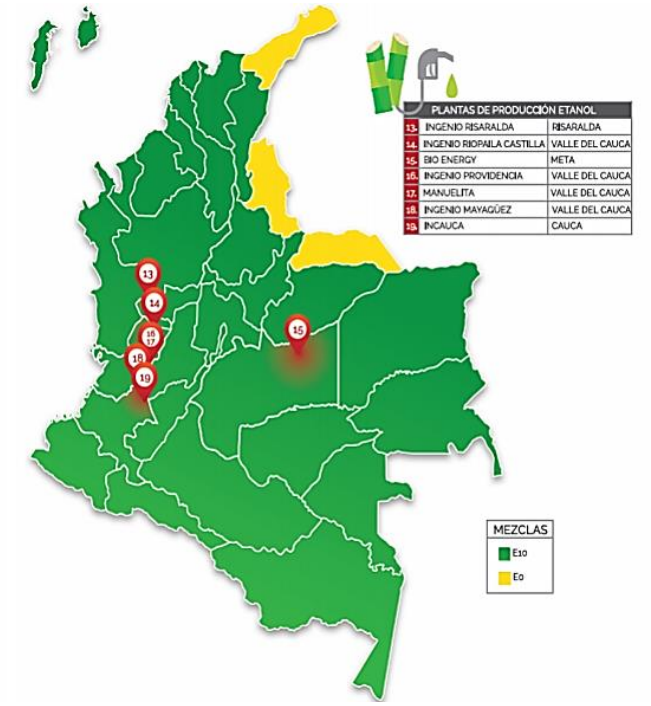


Figura 3. Distribución del porcentaje de mezcla de alcohol carburante (etanol) en el territorio nacional.

Nota: esta es la distribución actual, establecida desde marzo del 2019 y puede tener cambios en el futuro.

2.1.3 Materia prima utilizada en la producción de bioetanol

La materia prima en la producción de bioetanol es escogida teniendo en cuenta la viabilidad en cada región o país, como por ejemplo el porcentaje de caña de azúcar y la producción del cultivo plantado, entre otros aspectos.

La Figura 4 muestra la demanda nacional de etanol desde julio del 2018 hasta mayo del 2019. El punto máximo se presentó en enero del 2019 con una demanda nacional de 66,999,120 litros de etanol y hasta mayo del 2019 se observa una demanda nacional de 65,911,903 litros de etanol.

Teniendo en cuenta que el etanol puede ser producido por varias materias primas, en Colombia solo se utiliza caña de azúcar como cultivo de más productividad y rendimiento, con el cual se cubre el 75% de la demanda interna de etanol

carburante. Ésto se debe a la poca competitividad de las demás materias primas en términos económicos y productivos que se presentan en la región del Valle del Cauca y también por la logística e infraestructura ya implementadas en la región para la producción de azúcar [11].

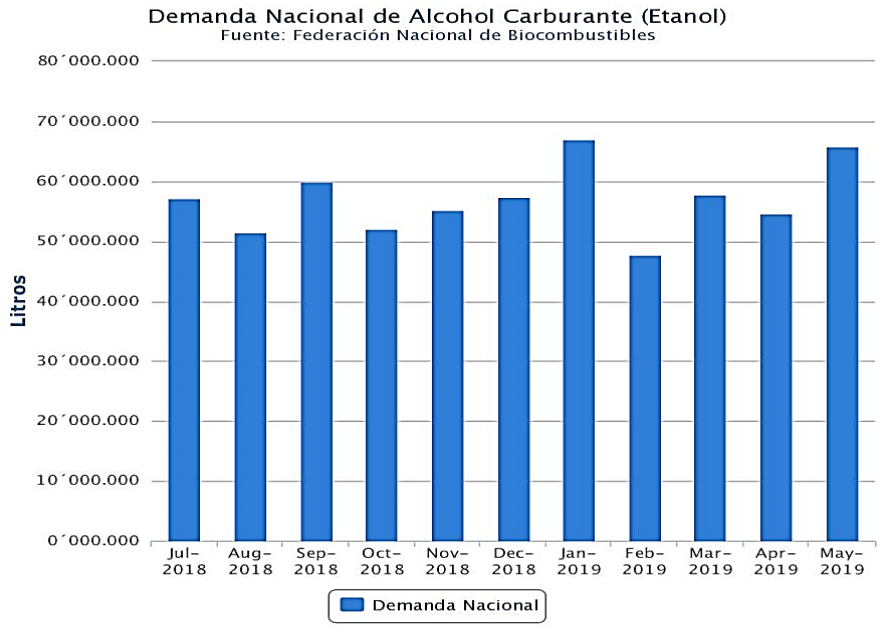


Figura 4. demanda nacional de alcohol carburante (etanol)

2.2 SINTESIS DEL BIOETANOL

El alcohol etílico más conocido como etanol es un biocombustible que se obtiene a partir de la fermentación de azucares provenientes de diversas materias primas tales como la caña de azúcar, el maíz, la remolacha o la yuca. En Colombia, el etanol o alcohol carburante es producido exclusivamente a partir del procesamiento de caña de azúcar [9].

El uso de etanol como combustible contribuye a reducir la contaminación puesto que disminuye las emisiones de CO_2 al reemplazar por ésta un porcentaje de gasolina que utilizan los automotores. Este tipo de mezclas permite aumentar la compresión en el motor, dando así un funcionamiento más regular, su

recalentamiento es menor y por ende se puede utilizar a un mayor número de revoluciones [9,16].

2.3 PRODUCCIÓN DE BIOETANOL DE CAÑA DE AZÚCAR

La Figura 5 muestra cada uno de los procesos que se llevan a cabo para la producción de etanol a partir de la caña de azúcar.

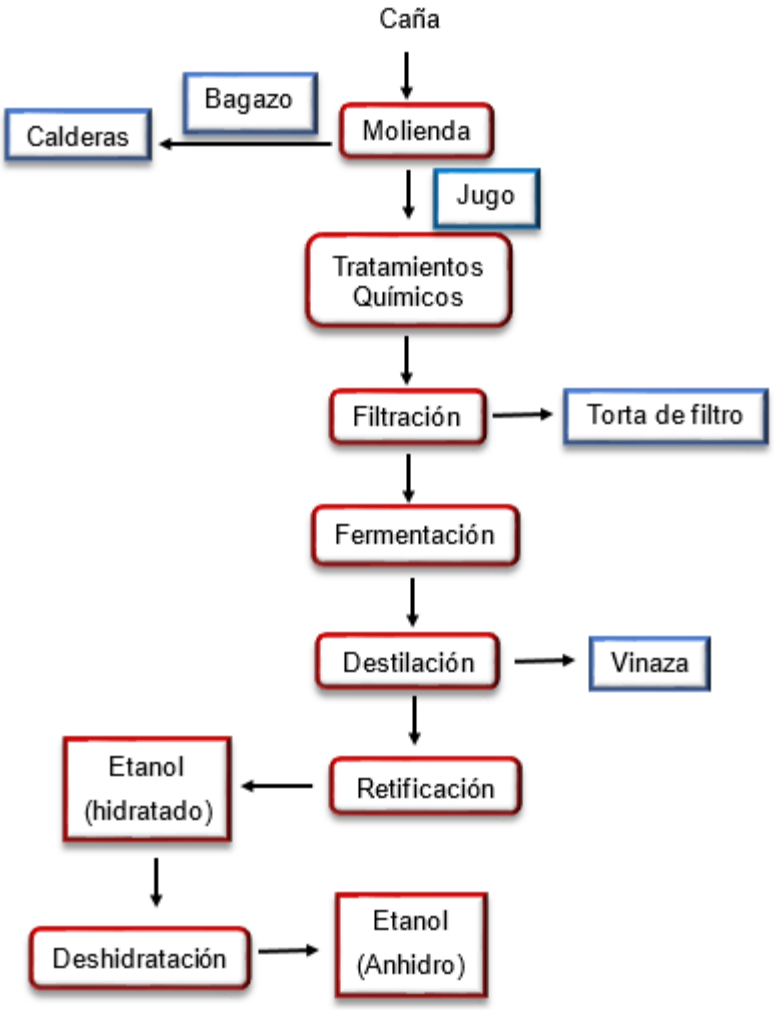


Figura 5. Esquema de proceso de extracción del Etanol en Colombia

2.4 ESPECTROSCOPIA DE INFRARROJO

2.4.1 Aspectos fundamentales

la espectroscopia molecular se basa en la interacción entre la radiación electromagnética y las moléculas. Dependiendo de la región del espectro en la que se trabaje y por tanto de la energía de la radiación utilizada (caracterizada por su longitud o número de onda), esta interacción será de diferente naturaleza: excitación de electrones, vibraciones moleculares y rotación moleculares. 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 de la espectroscopía infrarrojo (IR) es posible analizar muestras sólidas y líquidas. Se debe conocer el tipo de vibración que ocurre en los enlaces de la molécula, las cuales se muestran en la Figura 6. [17].

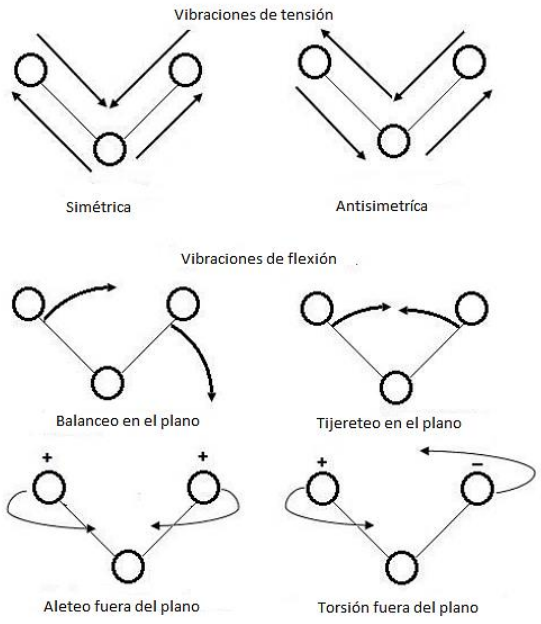


Figura 6. Tipos de vibraciones moleculares

Nota: (+) indica un movimiento del plano de la página hacia el observador; (-) indica un movimiento del plano de la página alejándose del observador.

2.4.2 Regiones Espectrales.

Aunque el espectro infrarrojo se extiende desde 10 a 14300 $\boldsymbol{cm^{-1}}$ desde un punto de vista funcional, éste se divide en tres zonas: IR lejano, donde se producen las absorciones debidas a cambios rotacionales, el IR medio (MIR), donde tienen lugar las vibraciones fundamentales y el IR cercano (NIR), donde se producen absorciones debidas a sobretonos y combinaciones de las bandas fundamentales, esto se puede apreciar en la Tabla 1 [18].

Tabla 1. División del espectro IR.

REGIÓN	TRANSICIÓN CARACTERÍSTICA	LONGITUD DE ONDA (NM)	NUMERO DE ONDA ($\boldsymbol{cm^{-1}}$)
Infrarrojo cercano (NIR)	Sobretono y combinaciones	700-2500	14300-4000
Infrarrojo medio	Vibraciones fundamentales	2500 - 5×10^4	4000-200
Infrarrojo lejano	rotaciones	5×10^4 - 10^6	200-10

2.4.3 Tipos de medidas en infrarrojo

Cuando la radiación incide en la muestra (Figura 7), ésta puede sufrir diferentes fenómenos: absorción, transmisión y reflexión. La intensidad de la luz transmitida a través de la muestra (P_T) es menor que la intensidad incidente (P_0). Una parte de

esta intensidad incidente se ha reflejado (P_R), mientras que otra parte ha sido absorbida por la sustancia (P_A).

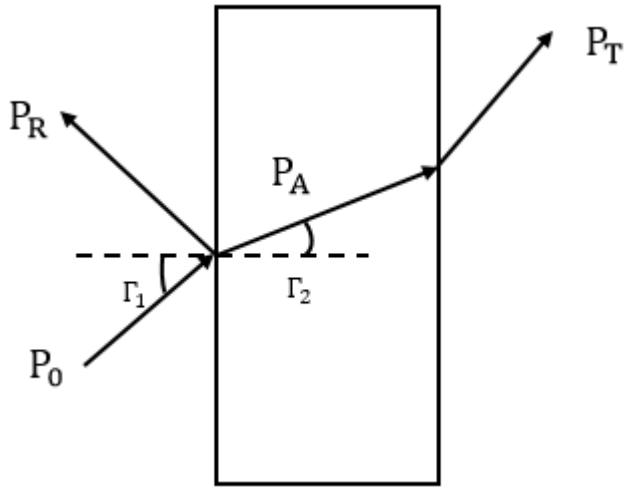


Figura 7. Fenómenos de absorción, transmisión y reflexión de la radiación electromagnética al interaccionar con la materia.

La medida más común en el infrarrojo es la que se basa en la absorción (o la intensidad transmitida), aunque también se han desarrollado espectroscopias basadas en el fenómeno de la reflexión como son la reflectancia total atenuada (ATR) y la reflectancia difusa [17].

2.4.4 Espectroscopia de infrarrojo cercano (NIR)

La espectroscopía de infrarrojo cercano es una técnica analítica que se basa en la absorción de radiación electromagnética en la región comprendida entre los 780 y 2500 nm (12820 – 4000 cm^{-1}). Se usa típicamente para medir de forma cuantitativa grupos funcionales orgánicos, especialmente O-H, N-H y C=O. Esta técnica permite analizar muestras de un modo no destructivo ni invasivo, no requiere de tratamientos previos y permite la obtención rápida de resultados analíticos.

La radiación NIR fue descubierta en 1800 por willian Herschel, un prestigioso astrónomo, quien, en su búsqueda por conocer si un color determinando del espectro obtenido al pasar la luz solar a través de un prisma se asociaba con la

temperatura de la luz solar, fue entonces cuando encontró que la máxima temperatura estaba más allá del rojo final del espectro visible [19].

El primer espectro del infrarrojo cercano fue medido en 1881 por Abney y Festing, usando una lámina fotográfica. Ellos no solo obtuvieron el primer espectro, sino que también sugirieron correctamente que la absorción se relacionaba con la composición de la muestra analizada [20].

Uno de los más importantes pioneros de la espectroscopia de reflectancia en el infrarrojo fue Willian Coblentz, quien en 1905 púbico un amplio estudio de compuestos cuyos espectros registró desde los 1000 a 16000 nm. El trabajo de Coblentz significó un avance para los investigadores. A partir de ese punto los investigadores fueron capaces de relacionar características de los grupos de átomos en las moléculas con absorciones especificas en el infrarrojo [20].

Los espectros en la región del NIR y MIR son originados a partir de la transferencia de radiación de las energías asociadas al movimiento de átomos en una molécula. Sin embargo, los espectros NIR son considerados más complejos comparados con los MIR. Este hecho se debe a la presencia de sobretonos y bandas de combinación que se pueden presentar en la técnica NIR. Otro motivo de complejidad son las contribuciónes de dos tipos de resonancia que pueden ocurrir en la región del NIR. Una de ellas es la resonancia de *Fermi* la cual puede ocurrir entre una absorción fundamental y un Sobretono cuando la diferencia de energía entre ellos es muy baja y, la otra de ellas es la resonancia llamada *Darling-Dennison* que puede promover la interacción entre dos sobretonos de nivel más alto de una molécula y la combinación de bandas, resultando el surgimiento de dos bandas en vez de una en la región de combinación [21].

Actualmente, ambas regiones (MIR y NIR) están siendo utilizadas para fines de determinación analítica, información cualitativa y cuantitativa de especies químicas,

Generalmente, para lograr interpretar los resultados obtenidos se necesitan de tratamientos matemáticos basados en herramientas quimiométricas, con el fin de extraer la mayor cantidad de información contenida en matrices complejas.

Los espectros en la región NIR corresponden casi que totalmente a sobretonos y combinaciones harmónicas de vibraciones fundamentales de enlaces N-H, C-H, O-H y S-H que aparecen normalmente en la región MIR [21].

La radiación NIR, al originar movimientos de distensión, rotación y vibración de los enlaces hidrógeno presentes en los constituyentes de la muestra, provoca patrones únicos de absorción en diversas longitudes de onda (espectros), que permiten, a través del análisis multivariado determinar la composición de la gasolina [22] [23].

En el análisis NIR, la calibración multivariada permite realizar las tres funciones básicas en la química analítica: separación, identificación y cuantificación de las propiedades físicas y químicas de las muestras, según sus espectros. Por lo tanto, la calibración multivariada es considerada la clave para el uso exitoso de la técnica NIR [24].

2.4.5 Algoritmo genético

Los algoritmos genéticos son procedimientos adaptivos que se usan para resolver problemas de búsqueda de la solución óptima a un problema. Están inspirados en la biología, y concretamente en la teoría de la evolución de las especies de Darwin.

Un investigador de la universidad de Michigan llamado John Holland estaba consciente de la importancia de la selección natural, y a fines de los 60s desarrollo una técnica que permitió incorporarla en un programa de computadora. Su objetivo era lograr que las computadoras “aprendieran por sí mismas”. A esta técnica que invento Holland se le llamo originalmente “planes reproductivos”. Pero se hizo popular bajo el nombre “algoritmo genético” tras la publicación de su libro en 1975, además los principios básicos de algoritmo genético se encuentran bien descritos

en varios textos – Gilbert (1989), Davis (1991), Michalewicz (1992). Reeves (1993) [25][26][27][28].

En síntesis, los algoritmos genéticos trabajan con una población de individuos, a los cuales se le asigna un valor relacionado con la bondad de dicha solución. En la naturaleza esto equivaldría al grado de efectividad de un organismo para así competir por unos determinados recursos. Ahora, cuanto mayor sea la adaptación de un individuo al problema, mayor será la probabilidad de que el mismo sea seleccionado para reproducirse, cruzando su material genético con otro individuo elegido de la misma forma.

Este cruce dará como resultado nuevos individuos los cuales compartirán alguna de las características de sus padres. Cuanto menor sea la adaptación de un individuo, menor será la probabilidad de que dicho individuo sea seleccionado para la reproducción, y por ende de que su material genético se propague en sucesivas generaciones. Es así como se produce una nueva población de posibles soluciones, la cual, reemplaza a la anterior y verifica la interesante propiedad de contener una mayor proporción de buenas características en comparación con la población anterior. De esta manera a lo largo de las generaciones las buenas características se propagan a través de la población. Esto favorece el cruce de los individuos mejor adaptados y van siendo exploradas las áreas más prometedoras del espacio de búsqueda [29].

2.5 QUIMIOMÉTRIA

Los instrumentos espectroscópicos de los laboratorios han llevado consigo diversas consecuencias. Una de las cuales, es la rápida adquisición de gran cantidad de datos en una única muestra. La obtención de dichos datos no es sinónimo de conocer lo que ocurre. Por ende, es necesario saber interpretarlos y colocarlos en un contexto adecuado para convertirlos en información útil para el usuario. La quimiometría es la disciplina que tiene esta finalidad [30].

Ésta es una disciplina metrológica que aplica conocimientos matemáticos, especialmente estadísticos, a procesos químicos para extraer de los datos experimentales la mayor cantidad posible de información y extender el conocimiento del sistema químico [31].

En la quimiometría aplicada a la química analítica el método multivariado es el de más utilidad. Su objetivo es predecir la propiedad de interés a partir de múltiples medidas instrumentales. Estos modelos son especialmente útiles para el análisis cuantitativo mediante técnicas espectroscópicas [32].

2.5.1 Pretratamiento de los datos

Las técnicas de pretratamiento de los datos pueden ser aplicadas tanto a las muestras como a las variables. Dentro de las técnicas de pretratamiento se encuentran suavizado, corrección de la línea base y derivadas.

- Suavizado (smoothing) espectral: se realiza con la finalidad de minimizar o eliminar el ruido espectral que puede interferir en los resultados de las respuestas analíticas de la variable en estudio. Se emplean filtros matemáticos como cálculos polinómicos (Savitzky-Golay) o por transformada de Fourier [33] [34].
- Correcciones de línea base: este método es comúnmente empleado en aplicaciones espectroscópicas donde la señal de algunas variables es debido solamente a la señal de fondo. Estas variables son utilizadas para determinar cuanta señal de fondo debe ser eliminada de las variables cercanas. El algoritmo matemático emplea un enfoque automático para así determinar qué puntos son los más probables para ser solo línea base; ésto lo hace ajustando una línea base a cada espectro y determinando que variables están claramente por arriba de la línea base (señal) [35].

  En realidad, existen muchos factores que pueden afectar la línea base, incluyendo la calidad del background del espectro, de la muestra, la forma en que se preparó la muestra, el tipo de accesorio empleado y el alineamiento

del divisor de haz. Debido a esto la línea base puede inclinarse, desplazarse o curvarse [36].

2.5.2 Construcción del modelo matemático

En el análisis NIR, la calibración permite realizar las tres funciones básicas en la química analítica: separación, identificación y cuantificación de las propiedades físicas y químicas de las muestras, según sus espectros. Por lo tanto, la calibración multivariada es considerada la clave para el uso exitoso de la técnica.

la calibración multivariada se encuentra relacionada con la propiedad de interés debido a la presencia de más de una variable utilizada para la predicción, el término multivariable es utilizado para el proceso.

La determinación de propiedades físicas y químicas de las muestras analizadas en el infrarrojo se da por métodos quimiométricos, más específicamente los métodos de calibración multivariable. Esto será realizado por métodos de modelos globales, tales como la regresión por componente principal (PCR) o mínimos cuadrados parciales (PLS)

2.5.3 Calibración multivariada

La calibración multivariada es probablemente el área de la quimiométria que ha atraído un mayor número de interés en aplicaciones de la espectrofotometría en la región del infrarrojo [35]. El propósito de la calibración multivariada es establecer una relación matemática cuantitativa entre la señal infrarroja obtenida mediante un espectrómetro y el parámetro fisicoquímico de interés previamente determinado por una técnica independiente [37]. La respuesta instrumental, para la construcción de los modelos, es representada en forma de matriz; mientras que las propiedades de interés, determinadas por una metodología patrón, son representadas por un vector [38]. La Figura 8 ilustra como una matriz de datos X de dimensión n x m, es decir, n objetos (espectros) y m variables (número de onda) que puede ser construida a partir de un vector de respuesta instrumental.

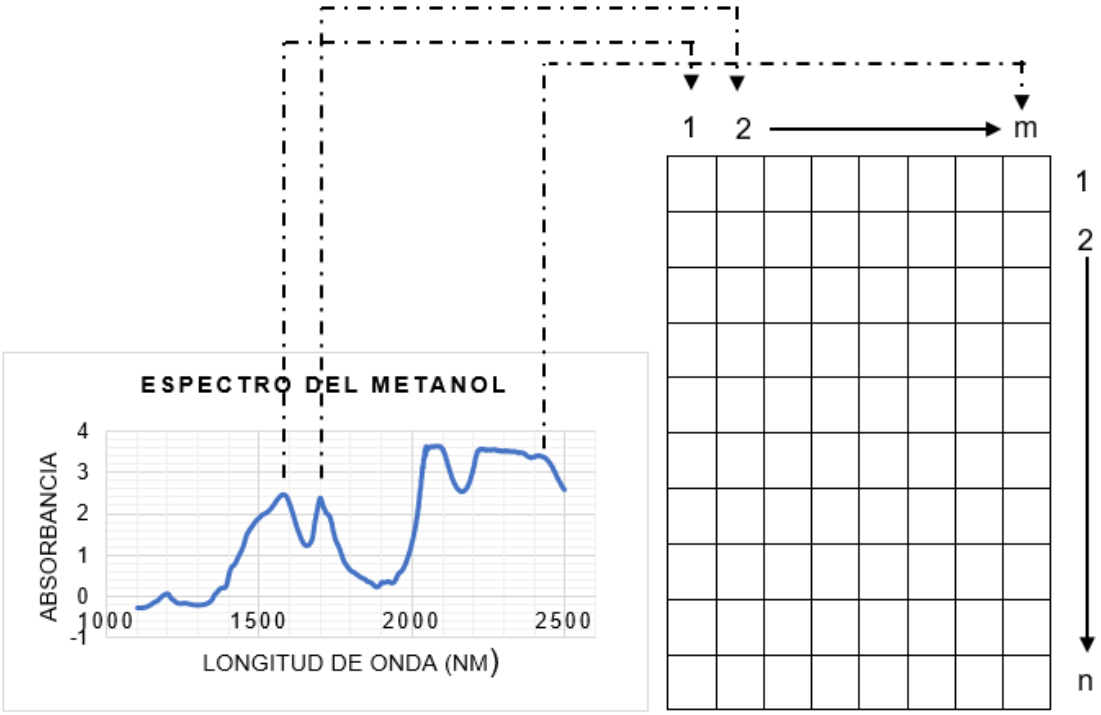


Figura 8. Representación de la construcción de la matriz de datos X.

Es importante destacar que, para el montaje de la matriz X, el programa Unscrambler® X10.4 no lee el espectro, pero si un archivo con números que forman el espectro; es decir, al obtener los espectros de cada una de las muestras corridas en el infrarrojo, estos deben ser exportados en un formato numérico para que puedan ser leídos por el software antes mencionado.

En el proceso de calibración multivariada basado en datos espectrales, tenemos para el presente modelo la relación entre valores de absorbancia. Un modelo de calibración, es entonces, una función matemática f(x) que relaciona dos grupos de variables, es decir, se relacionan los valores de Y en función de la regresión f (X):

$$Y = f(x) = Xb \qquad (Ecuación\ 1)$$

Donde:

Y = propiedad a inferir

X = Absorbancia a un número de onda.

b= constante

Esta etapa representa la calibración y por ende el conjunto de datos empleados para esta finalidad es llamado conjunto de calibración. Los parámetros del modelo son denominados coeficientes de regresión (b) determinados matemáticamente a partir de los datos experimentales [39].

2.5.4 Análisis en componente principal (PCA)

Esta es una técnica de reducción de variables que permite visualizar en un espacio de 2 o 3 dimensiones la similitud o diferencias que tienen un grupo de muestras entre sí.

La interpretación de los resultados obtenidos en el PCA para una clasificación es subjetiva y se lleva a cabo a partir de la representación de los scores de las muestras de un componente principal frente a los scores de otros componentes. Si existe una relación entre las muestras, en el gráfico de los scores, los puntos aparecerán agrupados; mientras que, si los puntos no se asemejan entre sí, estos aparecerán dispersos [40].

El PCA sintetiza un gran conjunto de datos, crea estructuras de interdependencia entre variables cuantitativas para crear unas nuevas variables que son función lineal de las originales y de las que podemos hacer una representación gráfica. El objetivo del análisis del componente principal es reducir la dimensión de un conjunto de p variables a un conjunto m de menor número de variables para mejorar la interpretación de los datos.

Las nuevas variables, los componentes principales determinan lo esencial de las variables originales, son una combinación lineal de ellas que además tienen unas propiedades interesantes:

1. Son ortogonales (cada componente representa una dirección del espacio de las variables originales).

2. La primera componente es la que más varianza contiene y la j-ésima tiene más varianza que la j+1 ésima...

El análisis del componente principal tiene como objetivo hallar combinaciones lineales de variables representativas de cierto fenómeno multidimensional, con la propiedad de que exhiban varianza máxima y que a la vez no estén correlacionadas entre sí.

La varianza de la componente es una expresión de la cantidad de información que lleva incorporada. Es decir, cuanto mayor sea su varianza, mayor será la cantidad de información incorporada de dicha componente.

La Figura 9 muestra de forma gráfica. Se tiene una matriz X_{ixj} (i filas por j columnas) la cual representan las i muestras de las j variables consideradas. El método de análisis de componente principal permite representar la variabilidad presente en X en unos pocos factores (componentes principales) que son combinaciones lineales de las variables originales [41] [42].

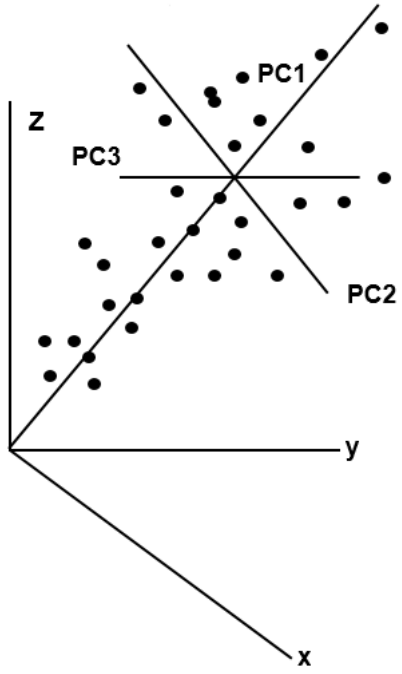


Figura 9. Representación gráfica de la descomposición en Componentes Principales de un conjunto de muestras definidas por tres variables.

El análisis de componentes principales proporciona una aproximación a la matriz X como producto de dos matrices: la matriz de scores T y la matriz de loadings P, que capturan la estructura de los datos de X. los scores capturan la estructura de las filas o lo que es lo mismo, las relaciones entre las muestras y los loadings retienen la relación existente entre las variables, como se muestra en la Figura 10.

$$X = TP^T + E \qquad (Ecuacion\ 2)$$

Donde E representa el error.

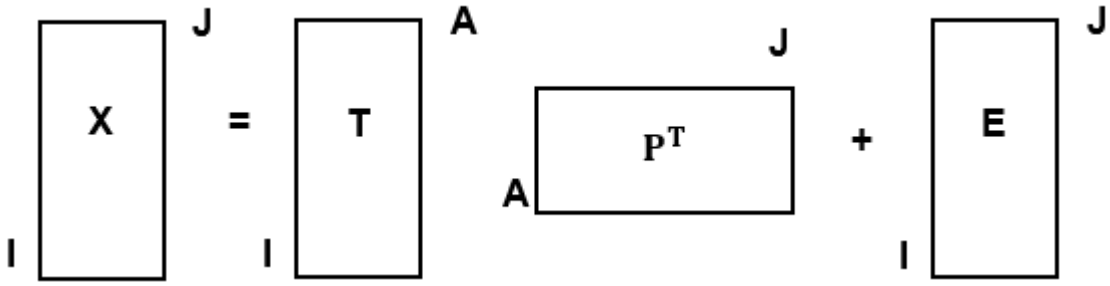


Figura 10. Notación Matricial de Componentes Principales.

2.5.5 Regresión de Mínimos cuadrados parciales (PLS)

La regresión de mínimos cuadrados parciales fue introducida por H. Wold (1975), para ser aplicada en ciencias económicas y sociales. Sin embargo, gracias a las contribuciones de su hijo Svante Wold, ha ganado popularidad en Quimiométria, en donde se analizan datos que se caracterizan por muchas variables predictoras, con problemas de multicolinealidad, y pocas unidades experimentales (observaciones o casos) en estudio. La idea motivadora de PLS fue heurística, por ello, algunas de sus propiedades son todavía desconocidas a pesar de los progresos alcanzados por Helland (1988), Hoskulsson (1988), Stone y Brooks (1990) y otros. La metodología PLS generaliza y combina característica del Análisis de Componentes Principales y Análisis de Regresión Múltiple. La demanda por esta metodología y la evidencia de que trabaja bien, van en aumento y así, la metodología PLS está siendo aplicada en muchas ramas de la ciencia de predictoras X. Éste calcula el

modelo de regresión estimado usando el vector de respuestas original y como predictoras, los componentes PLS. La reducción de la dimensionalidad puede ser aplicada directamente sobre los componentes ya que estos son ortogonales. El número de componentes necesarios para el análisis de regresión debe ser mucho menor que el número de predictoras [43].

La regresión por mínimos cuadrados parciales (PLS) es un método matemático que modela simultáneamente las matrices X y Y para encontrar un conjunto de variables latentes en X que mejor predicen las variables latentes en Y (Figura 11) [44] [45].

En el caso del análisis por componentes principales, estas nuevas variables en X y Y se pueden representar como un producto de matrices según se muestra en las Ecuaciones 3 y 4.

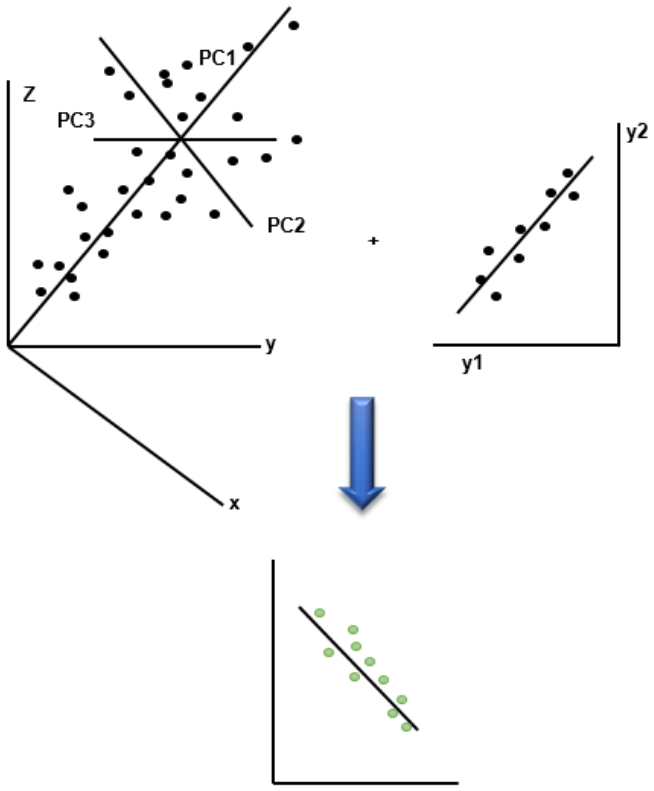


Figura 11. Descripción grafica del método de regresión PLS0.

$$X = TP^T + E = \sum t_a p_a^T + E \qquad (Ecuacion\ 3)$$

$$X = UQ^T + F = \sum u_a q_a^T + F \qquad (Ecuacion\ 4)$$

Donde:

T y U son las matrices de puntuación (scores) de X y Y respectivamente;

P y Q son las matrices de carga (loadings) de X y Y respectivamente;

E y F son los residuos.

3. METODOLOGÍA

3.1 Diseño de experimento

Se hizo uso del software estadístico JMP 12.2 64 bits para establecer el diseño de experimento para mezclas ternarias, el cual fue el diseño por restricción (Figura 12).

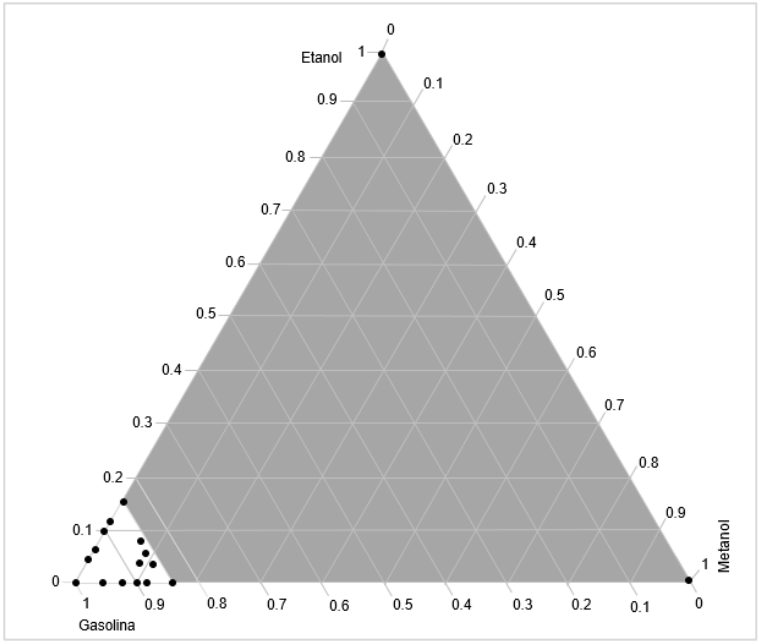


Figura 12. Diagrama ternario del diseño de mezclas por restricciones.

El intervalo de proporción expresado en %(V/V) para etanol y metanol fue de 0 a 15% y, el rango de porcentaje establecido para la gasolina fue de 85% a 100%.

3.2 Preparación de mezclas

Las mezclas de G/E/M se prepararon en el laboratorio del semillero de Investigación y Desarrollo en Bioproductos de la Universidad del Atlántico.

Para la preparación de las mismas se utilizó una micropipeta calibrada de 1000 μL (Ecopipete, $\pm$0,04) de volumen graduado y una balanza analítica (Bioprecisa, $\pm$0,0001).

Se prepararon diecisiete (17) mezclas G/E/M de 5 mL de acuerdo a las concentraciones definidas en el diseño de mezclas ternarias establecidas (Tabla 2). Con el objetivo de tener una mayor precisión en los datos se realizaron réplicas para cada una de ellas, para un total de treinta y cuatro (34) mezclas las cuales se etiquetaron con la letra A, por ejemplo, Metanol y Metanol A. Se utilizaron envases de vidrio, los cuales fueron adquiridos en la empresa Discordoba, ubicada en la ciudad de Barranquilla.

La muestra de gasolina fue suministrada por la empresa ECOPETROL S.A. y, tanto el metanol como el etanol fueron obtenidos en Merck Millipore.

Tabla 2. Proporción (%) y Volumen (***μ***L) de las muestras preparadas.

MUESTRAS	**PROPORCIÓN (%)**			**VOLUMEN (*μ*L)**		
	Metanol	Etanol	Gasolina	Metanol	Etanol	Gasolina
Metanol	100	0	0	5000	0	0
Etanol	0	100	0	0	5000	0
Gasolina	0	0	100	0	0	5000
E5	0	5	95	0	250	4750
E7	0	7	93	0	350	4650
E10	0	10	90	0	500	4500
E12	0	12	88	0	600	4400
E15	0	15	85	0	750	4250
M10	10	0	90	500	0	4500
M12	12	0	88	600	0	4400
M15	15	0	85	750	0	4250
M2.5E5	2.5	5	92.5	125	250	4625
M5	5	0	95	250	0	4750
M5E7.5	5	7.5	87.5	250	375	4375
M6E6	6	6	88	300	300	4400
M7	7	0	93	350	0	4650
M7.5E5	7.5	5	87.5	375	250	4375

3.3 Instrumentación

Los análisis NIR se realizaron en un espectrofotómetro FOSS NIRSystems 5000 para muestras líquidas (detector de transmitancia) el cual se encuentra disponible en el laboratorio de PROCAPS S.A. en la ciudad de Barranquilla.

Este equipo contiene un detector de diodo de PbS. La región espectral se encuentra en el intervalo 1100 a 2500 nm, con una resolución de 8 nm y una media correspondiente de 32 scan. Cada espectro se obtuvo utilizando viales de vidrio para IR de ocho (8) mm de diámetro. El tiempo utilizado para adquirir los espectros NIR fueron de 45 segundos.

Este equipo está conectado a una computadora equipada con el software ISIscan 2.21 el cual permite importar los datos del registro de los espectros obtenidos.

3.4 Pretratamientos de datos espectrales

A los datos adquiridos se le aplicó pretratamiento matemático con el fin de aumentar la relación señal/ruido. Éstos pretratamientos fueron: corrección de línea base la cual se evaluó mediante transformadas de desplazamiento (*baseline offset*); el suavizado de Savitzky-Golay con polinomio de orden dos (2) en los puntos de ventana 9,11 y 13. Además, se combinaron ambos tratamientos.

Para realizar estos pretatamientos matemáticos se hizo uso del software quimiométrico The Unscrambler® X 10.4. quien muestra los resultados de manera instantánea.

3.5 Algoritmo genético

Para conocer la zona espectral o las longitudes de onda donde se encuentra ubicada la mayor variabilidad de las muestras, fue necesario aplicar el algoritmo genético a la matriz de datos obtenidos con los pretratamientos seleccionados. Para ello, en un computador Lenovo Windows 10 con procesador Intel(R) Core(TM) i5-

7200U y sistema operativo de 64 bits se corrió tal algoritmo el cual le tomó un total de diez (10) días en mostrar las longitudes de onda seleccionadas para cada uno de los pretratamientos utilizados.

Para llevar a cabo este procedimiento, se utilizó el software estadístico R Project versión 3.6.0 junto con el paquete G.A versión 3.2 descargado de su librería, cumpliendo con las características mostradas en la Tabla 3.

Tabla 3. Condiciones del G.A.

PARÁMETROS	ESPECIFICACIONES
Tamaño de población	1000
Máximo número de generaciones	500
Ratio de mutación	0.005
Tipo de Algoritmo de cruce	Permutación

3.6 Construcción de los modelos de calibración multivariados

Después de conocer las longitudes de onda óptimas para el desarrollo de los modelos de calibración multivariados se procedió a seleccionar dos grupos de muestras en las datas definitivas. Para el primer grupo se eligieron un total de cincuenta y tres (53) muestras al azar para la construcción del modelo de calibración multivariado. Las quince (15) muestras restantes pertenecen al conjunto de validación cruzada. Ésto se muestra en la Figura 13.

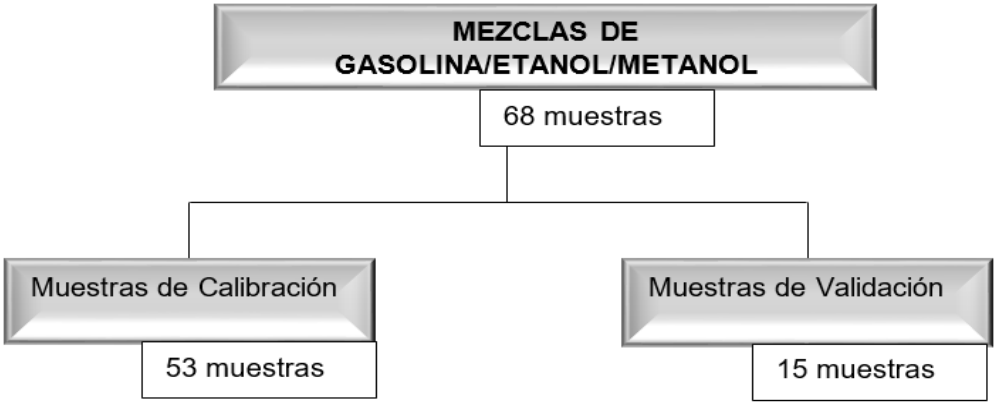


Figura 13. Distribución de las mezclas preparadas para el desarrollo de los modelos de calibración y validación.

Los cálculos estadísticos y gráficos correspondientes se realizaron con el software Quimiométrico Unscrambler® X 10.4. Se realizó un análisis de componentes principales para la matriz completa sin pretratamiento matemático. Además, se utilizó el PLS para la construcción de los modelos de calibración con las longitudes de onda definidas en el algoritmo genético para cada uno de los pretratamientos establecidos.

La capacidad predictiva del modelo es evaluada utilizando parámetros estadísticos como el Error Medio Cuadrático de Predicción (RMSEP) y el coeficiente de correlación entre la concentración real de alcohol en las mezclas y la concentración predicha en la validación (Q^2).

4. RESULTADOS Y DISCUSIÓN

4.1 Análisis de espectros NIR

La Figura 14 muestra tres (3) espectros NIR correspondientes a los componentes puros de las mezclas (gasolina, metanol y etanol). Es posible observar, por un lado, las similitudes que éstos presentan en el rango espectral de 1100 a 1330 nm y de 1680 a 1780 nm, debido a que en sus estructuras contienen enlaces del mismo tipo. Por otro lado, se nota una diferencia espectral en el rango de longitud de onda de 1401 a 1640 nm y de 1870 a 2000 nm. El espectro del etanol se superpone en el espectro del metanol en el rango espectral de 2010 a 2146 nm.

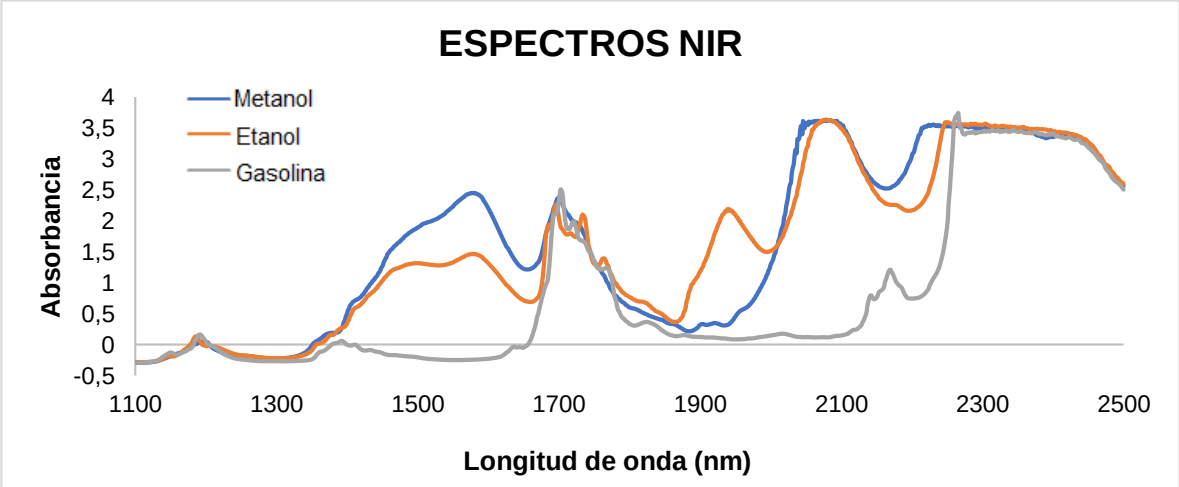


Figura 14. Espectros NIR de los componentes puros de la mezcla.

La Figura 15 muestra las estructuras moleculares del metanol, etanol y gasolina, donde se puede observar que cada uno de los componentes contienen enlaces de tipo $C-H$. Una diferencia existente entre estos compuestos en la presencia del enlace $O-H$ característicos de los alcoholes, mientras que la gasolina solo contiene enlaces $C-H$. Además, el metanol y el etanol difieren en un carbono $(-CH_2)$.

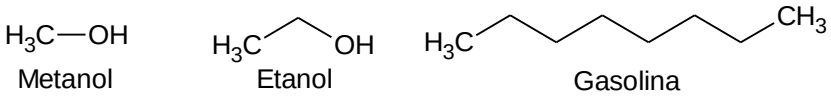


Figura 15. Estructura Molecular del Metanol, Etanol y Gasolina.

Es posible, relacionar las bandas presentes en los espectros NIR (Figura 14) con las estructuras químicas correspondientes a la gasolina, el metanol y el etanol (Figura 15), obteniendo así una correlación espectro-estructura.

En el rango espectral de 1115 a 1360 nm se observa una superposición de los tres espectros, donde se nota una banda de absorción entre 1120 a 1270 nm la cual pertenece al segundo sobretono de los modos de estiramiento del enlace $C-H$ quien se encuentra tanto en la gasolina como en los alcoholes. Además, un grupo de bandas se observa en la región espectral de 1688 a 1781 nm las cuales son atribuidas al primer sobretono de los modos de estiramiento del CH_3.

Otra banda se observa en la región de longitud de onda de 1350 y 1670 nm la cual resulta de una superposición de dos bandas de absorción. La primera de ellas está ubicada entre 1350 y 1550 nm y se relaciona con el primer sobretono de la banda de estiramiento de los enlaces $C-H+C-H$ y $C-H+C-C$ y, la segunda de ellas, se encuentra entre 1400 y 1670 nm la cual es la región asignada al primer sobretono de los modos de estiramiento del enlace $O-H$ que se encuentra en los alcoholes [1].

El etanol presenta una banda en la región espectral de 1900 a 2000 nm correspondiente al primer sobretono de los modos de estiramiento del CH_2, el cual se encuentra unido al OH en este compuesto. Las bandas de combinación pertenecientes a los enlaces $C-H+C-H$ tanto del metanol como del etanol se observan en la región espectral de 2000 a 2200 nm [46].

La Figura 16 muestra la información espectral obtenido para sesenta y cuatro (64) espectros NIR. La mayor variación espectral se observa mayormente en el rango de longitud de onda de 1389 a 2225 nm, el cual fue seleccionado para la construcción de los modelos de calibración multivariados.

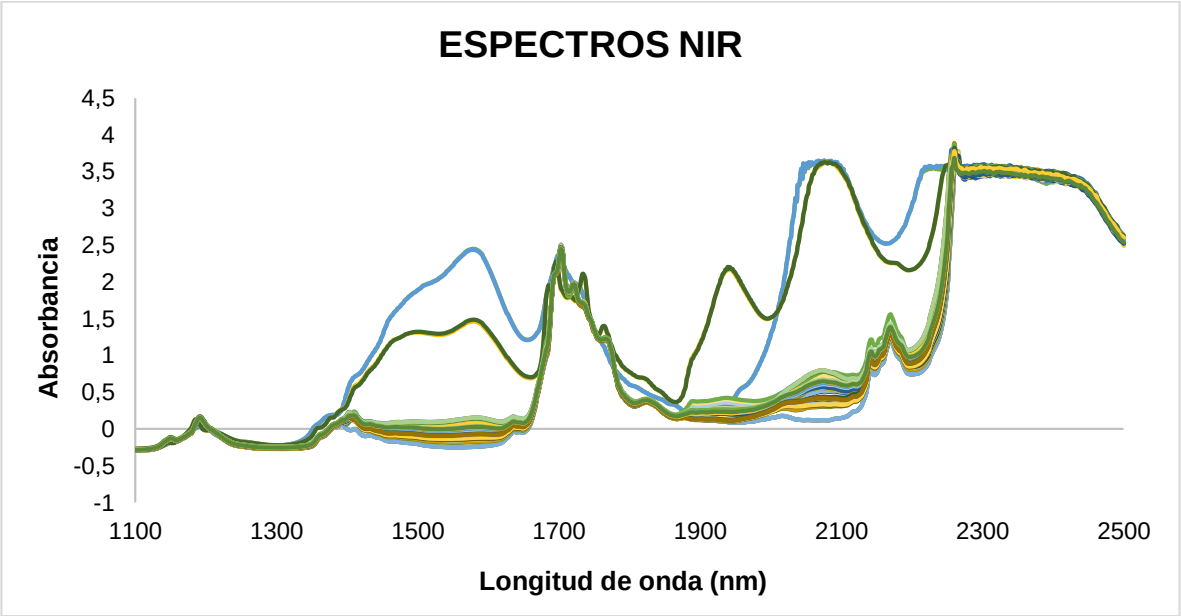


Figura 16. Espectros de las 16 mezclas de G/M/E preparadas.

4.2 Análisis de Componentes Principales (PCA)

Antes de iniciar la construcción de los modelos de calibración multivariados, se inició con el análisis de componentes principales (PCA por sus siglas en inglés) con la finalidad de determinar la principal fuente de variación en las mezclas. Este análisis se divide en cuatro (4) gráficos. El primero de ellos muestra los scores, quienes son utilizados para determinar el grado de similitud entre las muestras. En segundo lugar, los loadings nos permiten analizar la influencia de las distintas longitudes de onda. En tercer lugar, está la varianza explicada que se utiliza para identificar la mayor variación en X y, por último, encontramos el influence el cual identifica las muestras outliers.

La Figura 17 muestra el gráfico de los scores de los dos primeros componentes principales (PCs): PC1 Vs PC2, el cual explica el porcentaje más alto de la variación de los datos. En éste se observa como las muestras están ubicadas en un espacio bidimensional cuyos factores principales que explican la mayor variabilidad de los

espectros son los ejes. Además, la distribución de los mismos depende de las variaciones que haya entre ellos.

Se observa una figura triangular indicando el estudio de una mezcla ternaria. Cada uno de los extremos representan los espectros correspondientes a las muestras puras (gasolina, metanol y etanol). Un mayor agrupamiento se nota en una de las esquinas de éste, indicando que las muestras tienen un mayor contenido de gasolina. En el lado superior del triángulo se localizan las muestras de solo gasolina y etanol, de igual manera, en el lado inferior del mismo se ubican aquellas muestras con contenido solo de gasolina y metanol y, en el centro se observan las muestras con los tres componentes en estudio.

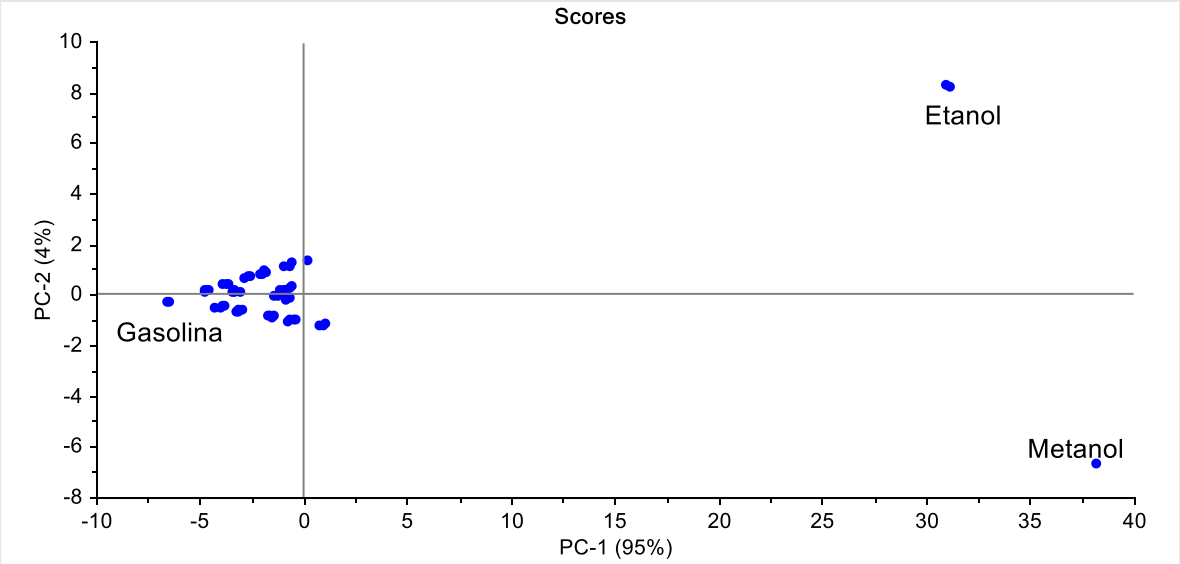


Figura 17. Scores de las mezclas G/M/E preparadas.

Las Figuras 18, 19 y 20 presentan los gráficos de los loadings. Los picos más pronunciados con respecto a las longitudes de onda del primer componente (PC1) representan una mayor influencia (variabilidad del 95%) para el análisis de las muestras, lo que en el PC1 se encuentra la mayor varianza, es decir, que éste contiene una mayor información acerca de las variables. Igualmente se observa que la mayor variabilidad se encuentra en el rango espectral de 1389 a 2500 nm. Esta información se puede observar en la Figura 18.

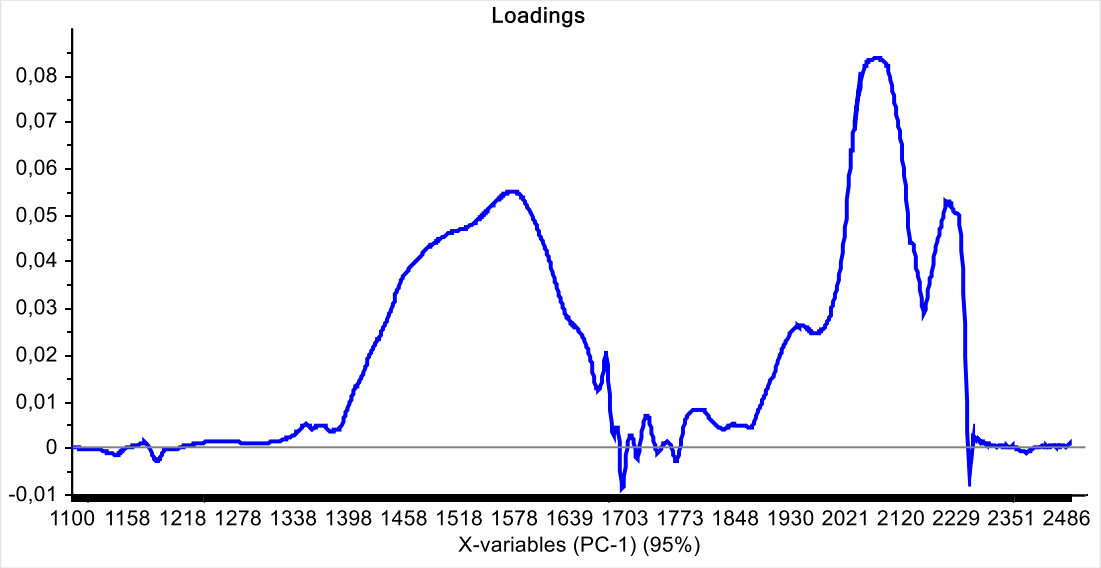


Figura 18. Loadings del PC1 de las mezclas G/M/E preparadas.

El 5% restante perteneciente a la variabilidad total se encuentra entre el segundo y tercer componente (PC2 y PC3) como se muestra en las Figuras 19 y 20. Teniendo en cuenta estos tres PCs se obtiene un total del 100% de la variabilidad.

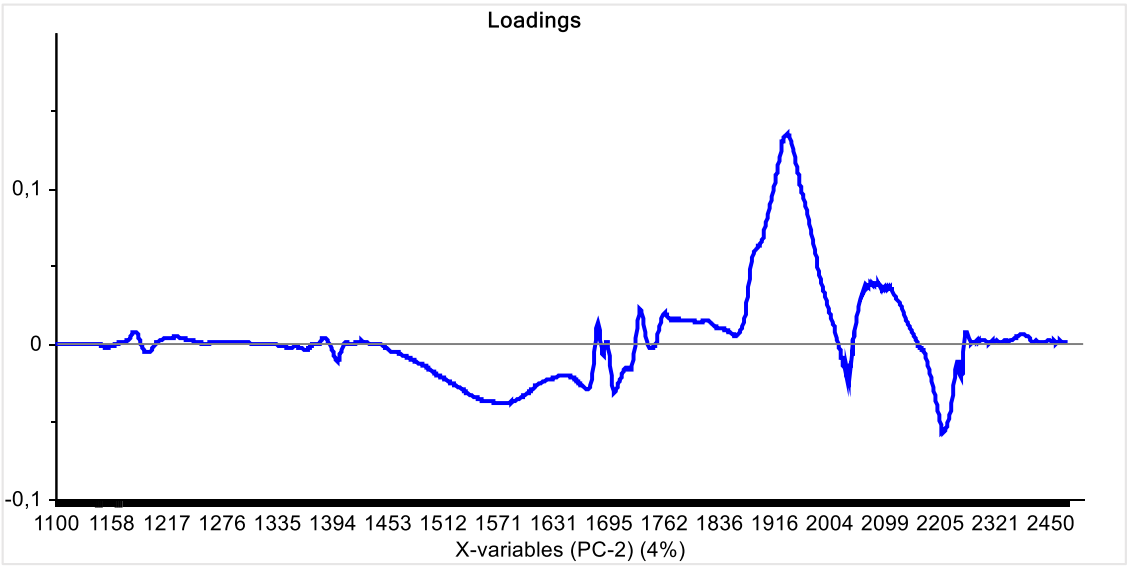


Figura 19. Loadings del PC2 de las mezclas G/M/E preparadas.

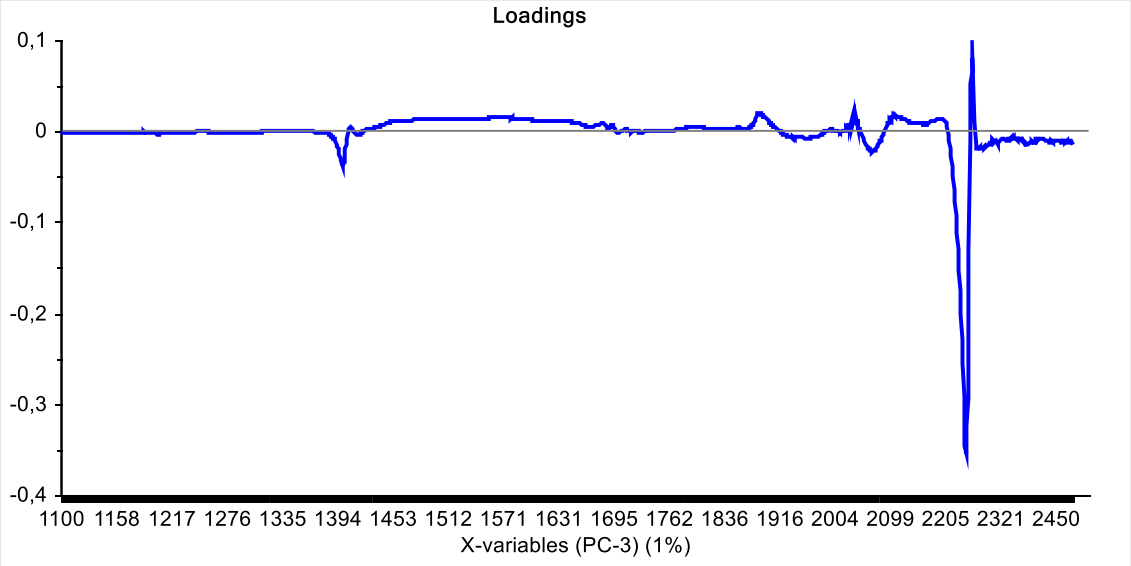


Figura 20. Loadings del PC3 de las mezclas G/M/E preparadas.

Lo anterior, se confirma en la Figura 21 el cual muestra la varianza explicada, éste indica la cantidad de variación en los datos que describen cada componente. Por ejemplo, en el primer componente se observa la mayor variabilidad, verificando de forma correcta la información anterior.

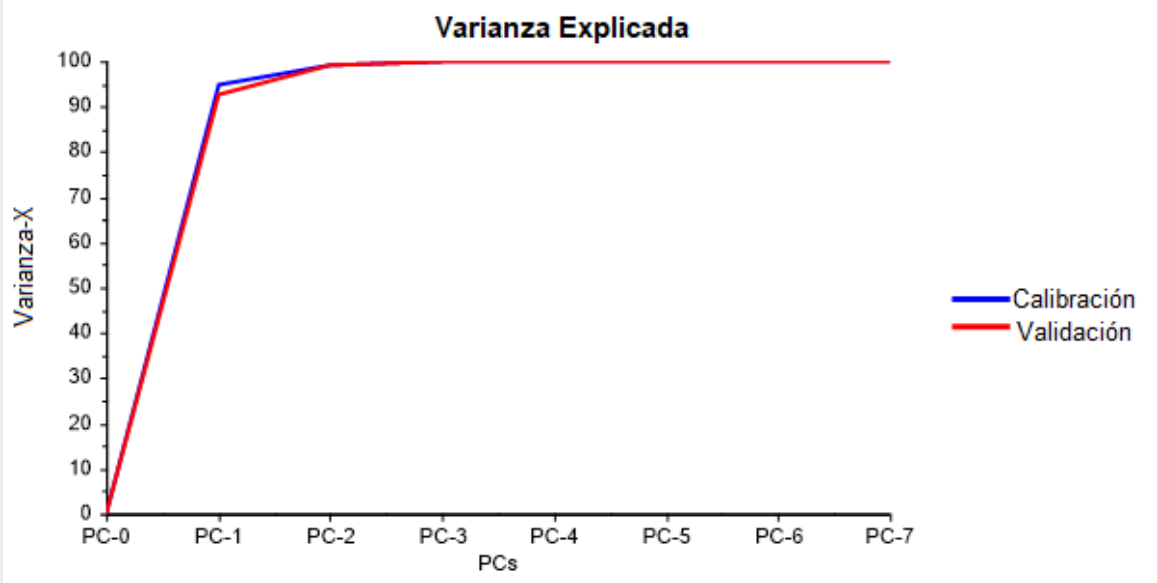


Figura 21. Varianza explicada con respecto a las variables X Vs PCs.

Varias causas posibles pueden contribuir a la aparición de muestras atípicas o también conocidas como outliers. La mayoría de ellas provienen comúnmente de errores en la preparación de las muestras o problemas que ocurren en el momento de la adquisición de espectros. Debido a que estas muestras influyen de forma negativa en el desarrollo de los modelos de calibración deben ser identificadas. Para ello, se aplica a los scores el test de outliers. Conociendo que las muestras ubicadas fuera de ésta serán consideradas como atípicas (ver Figura 22).

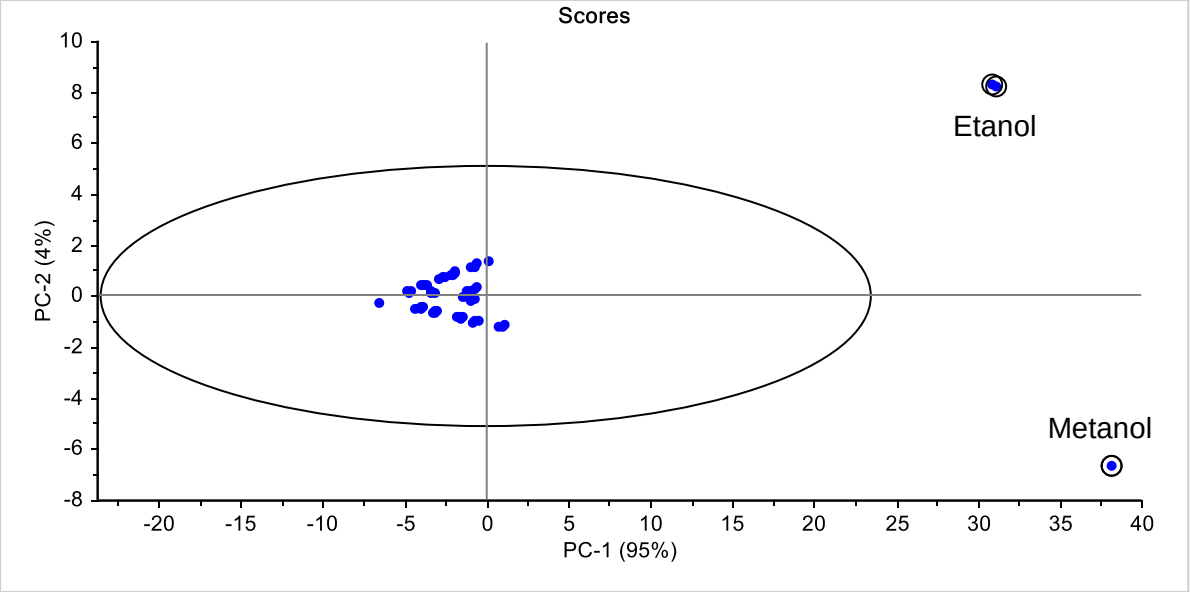


Figura 22. Elipse T^2.

Este test sugiere que las muestras outliers son el metanol y el etanol en su forma pura (al 100%). Sin embargo, la influencia de éstas a la hora de desarrollar el modelo de calibración no es alta, es decir, tienen una varianza residual baja, debido a ésto,

tanto el metanol como el etanol son consideradas muestras influyentes en la construcción del modelo, y no necesariamente outliers.

Es posible verificar la información anterior en el gráfico de los F-residual (Figura 23)

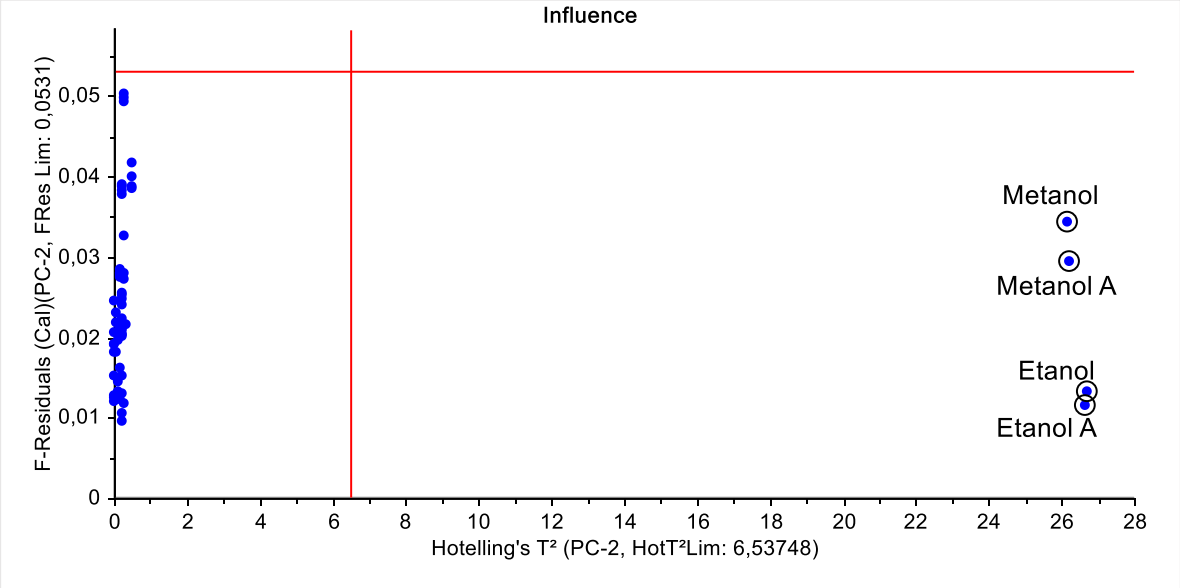


Figura 23. Gráficos de F-residuales, Test de Hotelling's.

Por un lado, el modelo toma como muestras outliers aquellas con alta varianza residual, es decir, que se encuentran por encima de la línea horizontal la cual representa el límite de residuos. Por otro lado, el modelo describe bien aquellas muestras que tienen una alta influencia, es decir, las que se encuentran del lado izquierdo del gráfico. Están bien descritos en el sentido de que los puntajes de la muestra pueden tener tanto, valores altos como bajos para algunos componentes en comparación con el resto de las muestras. Estas muestras tienen una alta influencia en el modelo de calibración.

4.3 Algoritmo Genético.

El algoritmo genético se corrió en el software estadístico R Project versión 3.6.0 con cada una de las matrices obtenidas, sin pretratamiento (Figura 24), suavizado de Savitzky-Golay (Figura 25), corrección de línea base (Figura 26) y suavizado-línea

base (Figura 27) en las cuales el G.A seleccionó las longitudes de onda donde se encuentra la mayor variabilidad de las muestras de manera más exacta, razón por la cual, los modelos se construyeron con las longitudes de onda escogidas.

A continuación, se presentan cuatro (4) espectros NIR del metanol mostrando de manera gráfica y aproximada las longitudes de onda seleccionadas por el algoritmo genético (líneas amarillas) para cada una de los pretratamientos matemáticos mencionados.

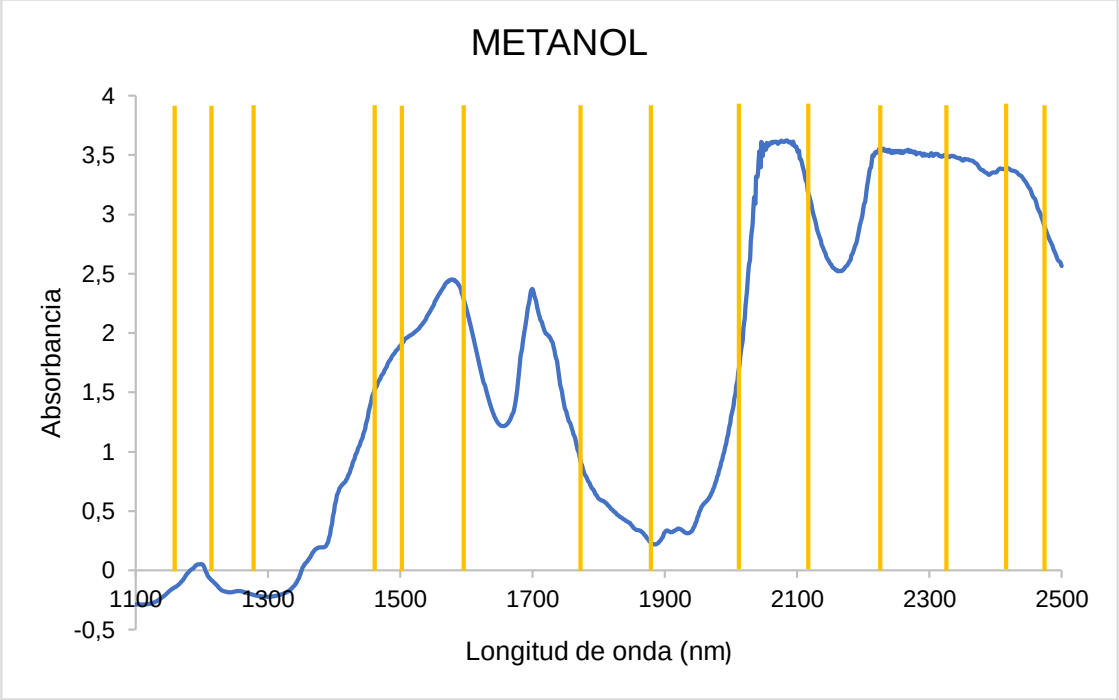


Figura 24. Longitudes de onda seleccionas por el G.A para matriz sin pretratamiento matemático.

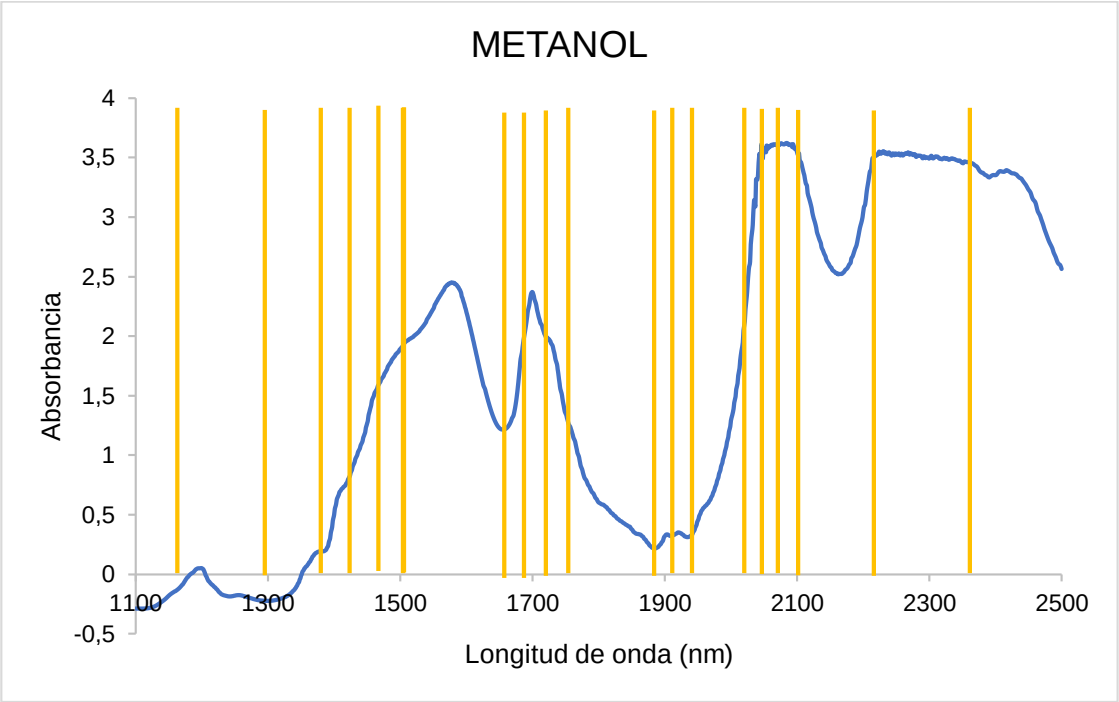


Figura 25. Longitudes de onda seleccionada por el G.A para matriz con suavizado

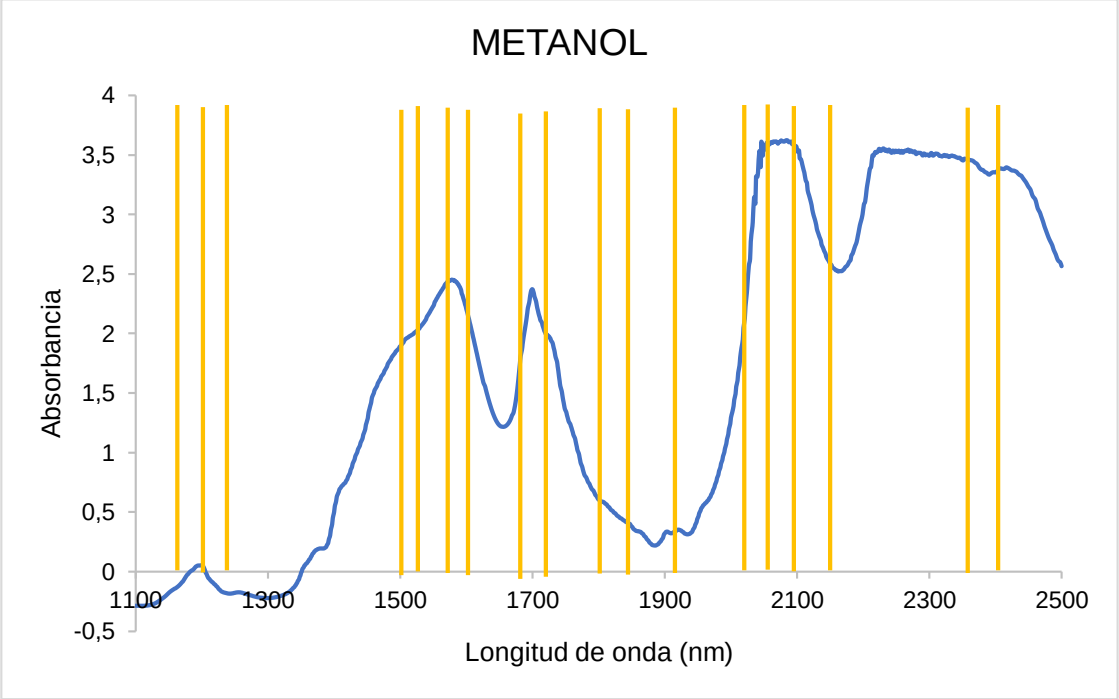


Figura 26. Longitudes de onda seleccionas por el G.A para matriz con corrección de línea base.

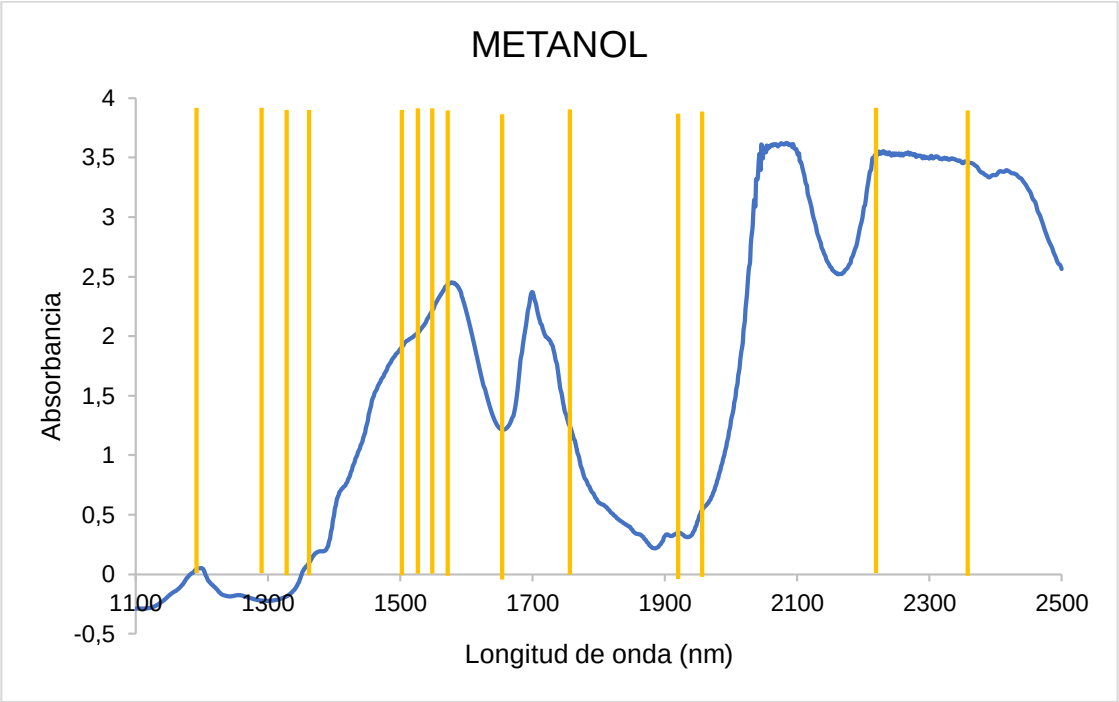


Figura 27. Longitudes de onda seleccionas por el G.A para matriz con suavizado-Corrección de línea base.

El algoritmo genético seleccionó las longitudes de onda para cada una de las matrices obtenidas de forma alterna, es decir, en rangos discontinuos los cuales son distintos para cada uno de los pretratamientos matemáticos establecidos.

En el rango espectral de 2000 a 2500 nm aproximadamente, se observa un factor común, el algoritmo genético no realiza selección alguna en esta zona debido a que allí se encuentra la señal de ruido, por lo cual no se obtiene información útil sobre la variabilidad de las muestras.

Las longitudes de onda seleccionadas para la matriz sin pretratamiento (Figura 24) y para la matriz con los dos pretratamientos matemáticos (Figura 27) son muy similares, en ambas se presentan rangos discontinuos amplios, es decir, las longitudes de onda seleccionadas se encuentran separadas en intervalos relativamente grandes. Sin embargo, cuando se aplican ambos pretratamientos (Figura 27) la selección en el rango de 1300 a 1600 nm contiene un mayor número de longitudes onda.

Ésto mismo ocurre con la matriz a la cual se le aplicó suavizado de Savitzky-Golay (Figura 25). No obstante, las longitudes de onda seleccionadas a lo largo del rango espectral fueron elegidas en pequeños rangos.
Finalmente, para la matriz con corrección de línea base (Figura 26) la selección fue mayor, abarcando un amplio número de longitudes de onda lo cual permitió comprender un rango de variabilidad más completo.

4.4 Construcción de modelos de calibración multivariable: PLS

Las Figuras de la 28 a la 35 hacen referencia a la varianza explicada para cada Y-variables con números diferentes de componentes principales en cada uno de los modelos diseñados con y sin algoritmo genético. cabe resaltar que cada grafica en donde se ve expresada la varianza explicada contiene datos de las concentración y longitudes de onda para cada una de los componentes, El intervalo de concentración expresado en %(V/V) para las variables etanol y metanol fue de 0 a 15% y, el rango de porcentaje establecido para la variable gasolina fue de 85% a 100%. En la construcción de los modelos de calibración multivariable Se puede observar que cada una de las variables se encuentran bien descritas por cada uno de los modelos desarrollados.

La línea azul en los gráficos de varianza explicada muestra el ajuste de los datos de calibración para el modelo, mientras que la línea roja hace referencia a la varianza de validación, este se calcula con los datos que no fueron tenidos en cuenta para la construcción del modelo.

Las variaciones explicadas se calculan para modelos que incluyen diferentes números de componentes principales.

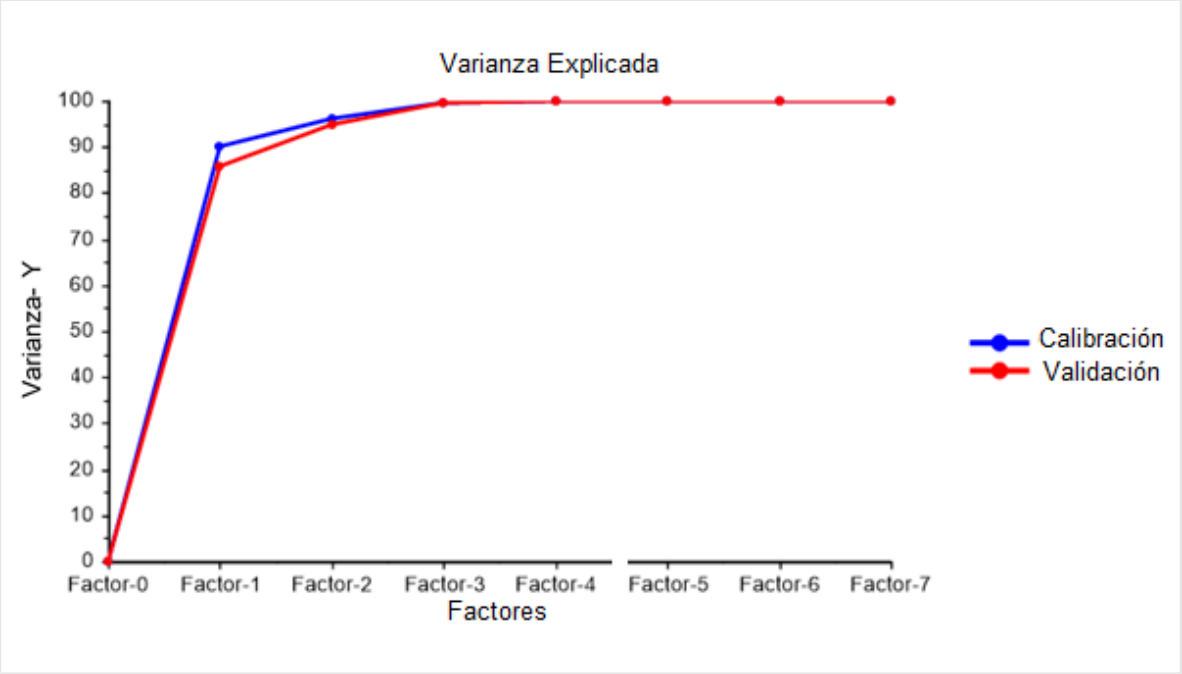


Figura 28. Gráfico de varianza en Y vs número de factores del PLS matriz sin pretratamiento.

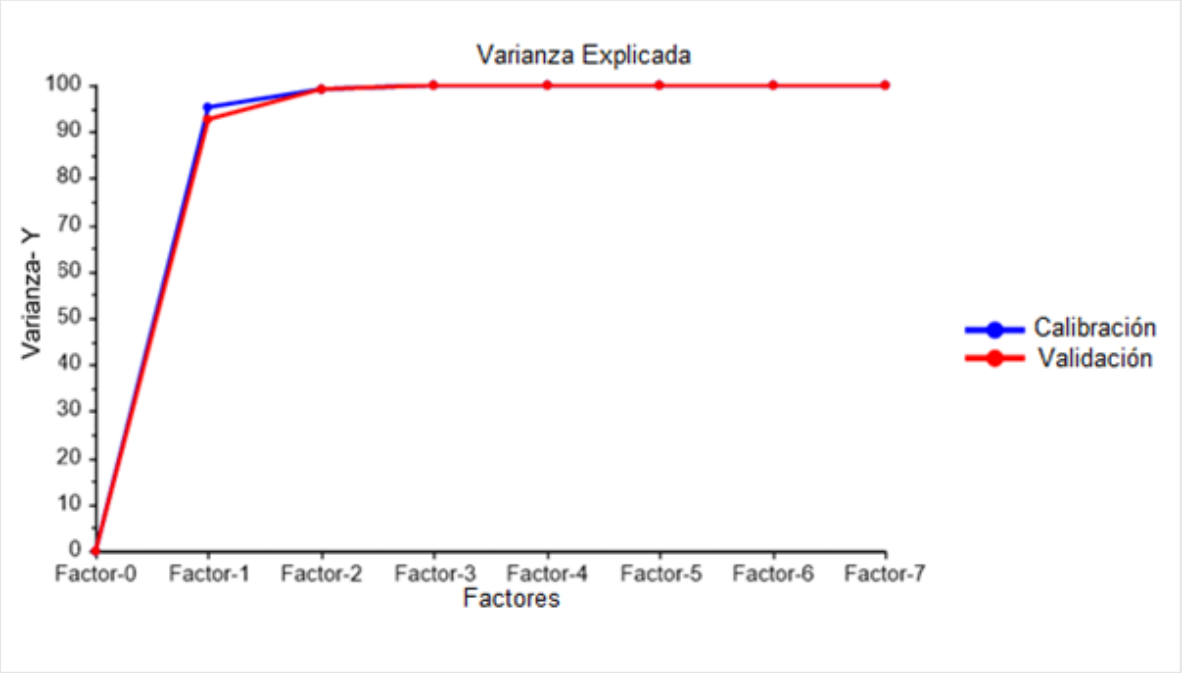


Figura 29. Gráfico de varianza en Y vs número de factores del PLS matriz con algoritmo genético.

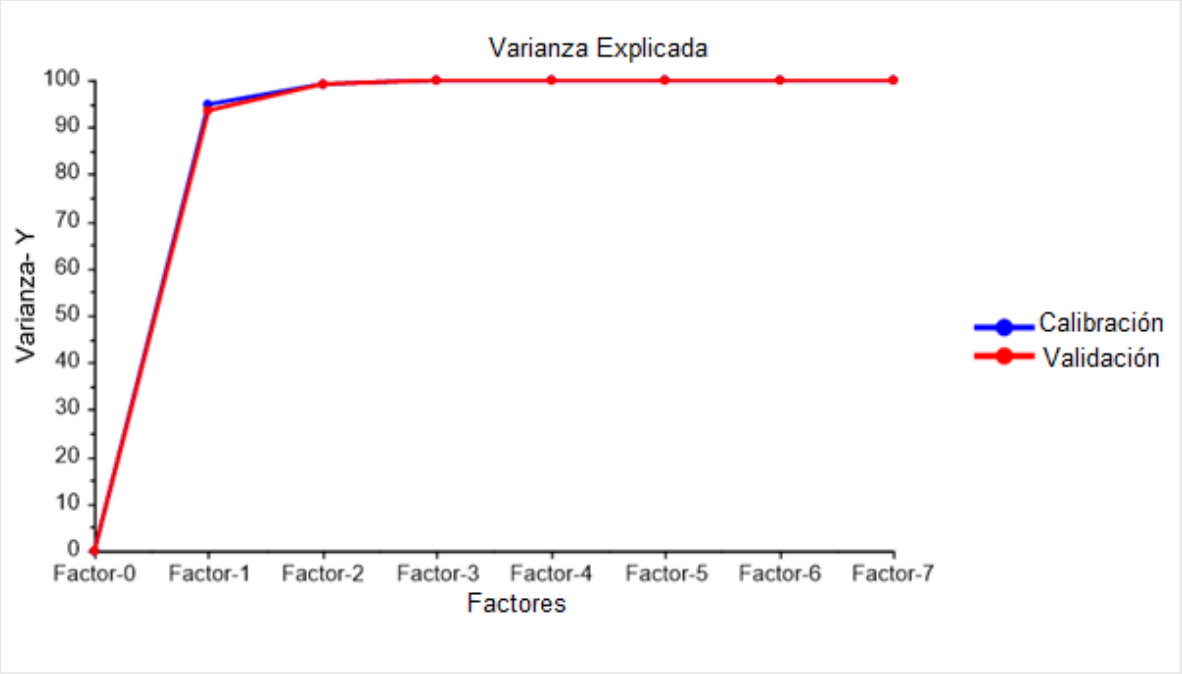


Figura 30. Gráfico de varianza en Y vs número de factores del PLS matriz Smoothing SGolay.

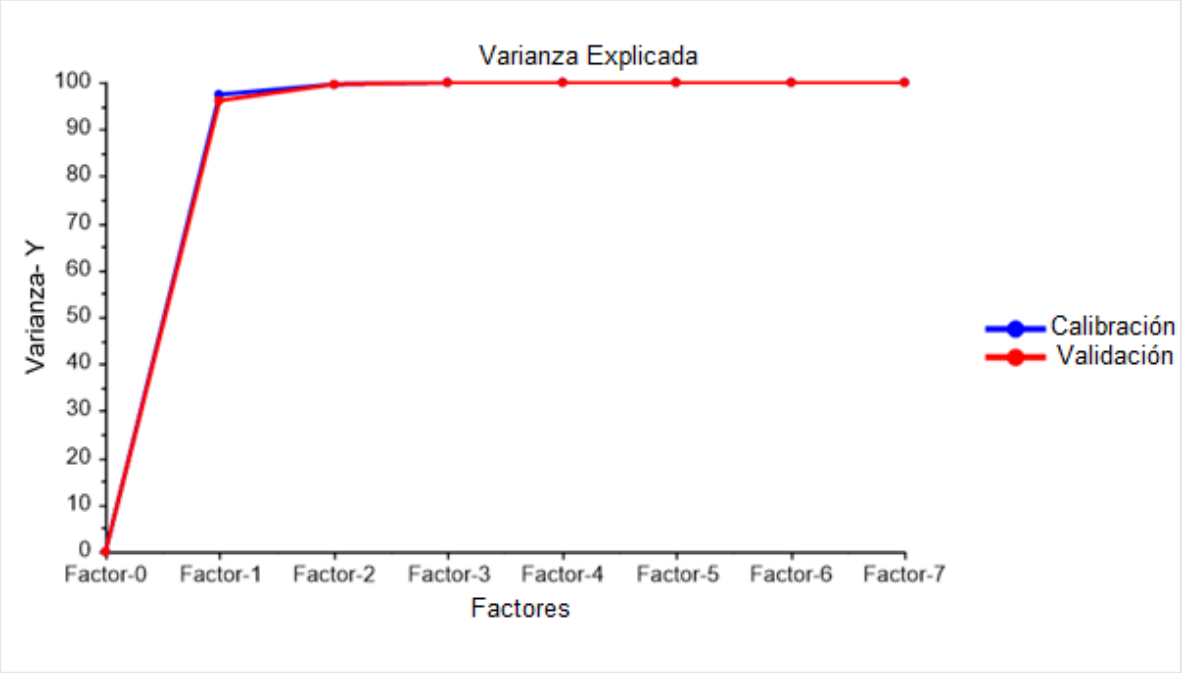


Figura 31. Gráfico de varianza en Y vs número de factores del PLS; algoritmo genético con Smoothing SGolay.

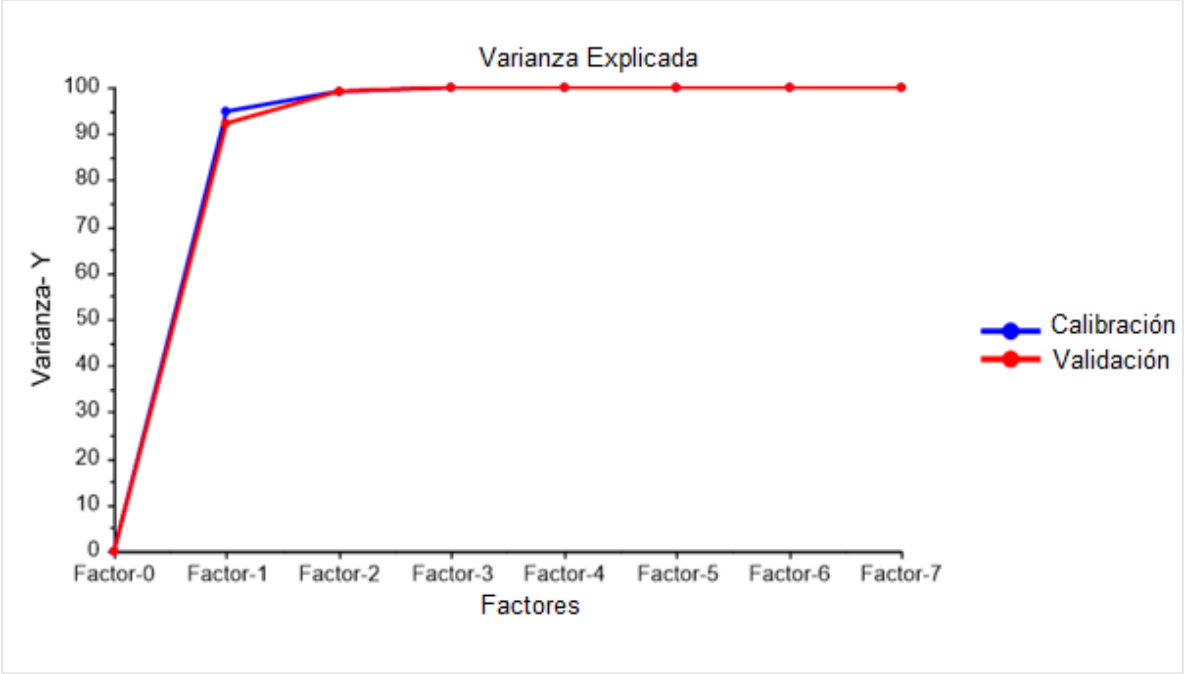


Figura 32. Gráfico de varianza en Y vs número de factores del PLS con Corrección de Línea Base.

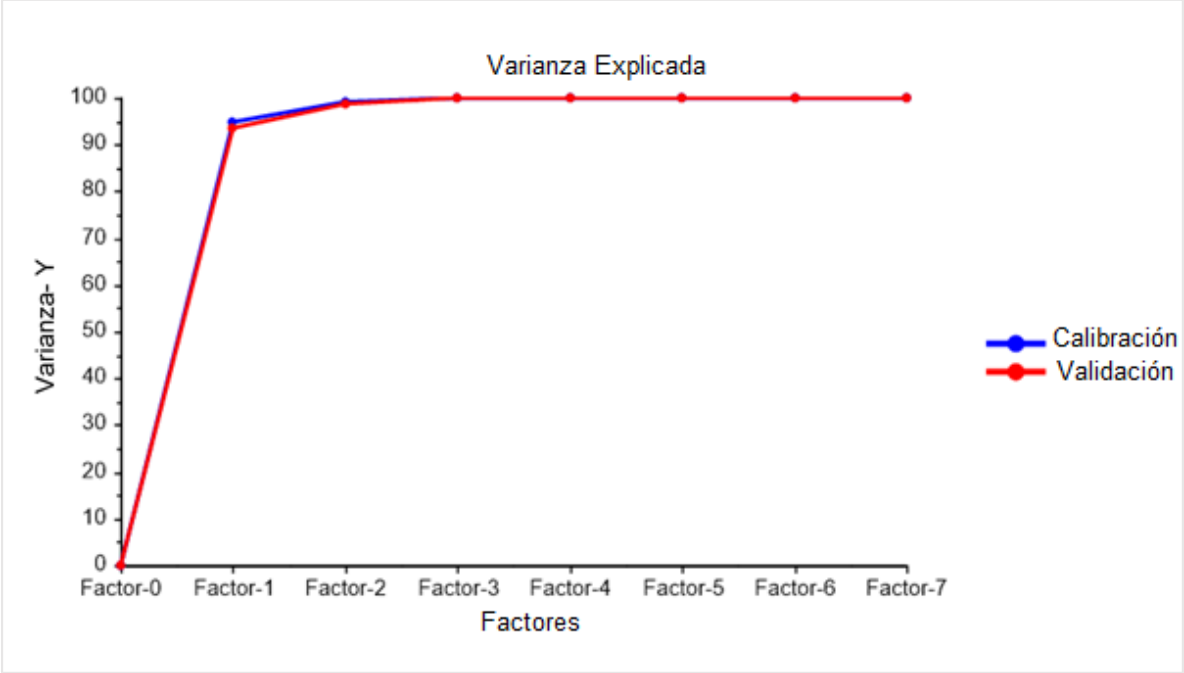


Figura 33. Gráfico de varianza en Y vs número de factores del PLS; algoritmo genético con corrección de línea base.

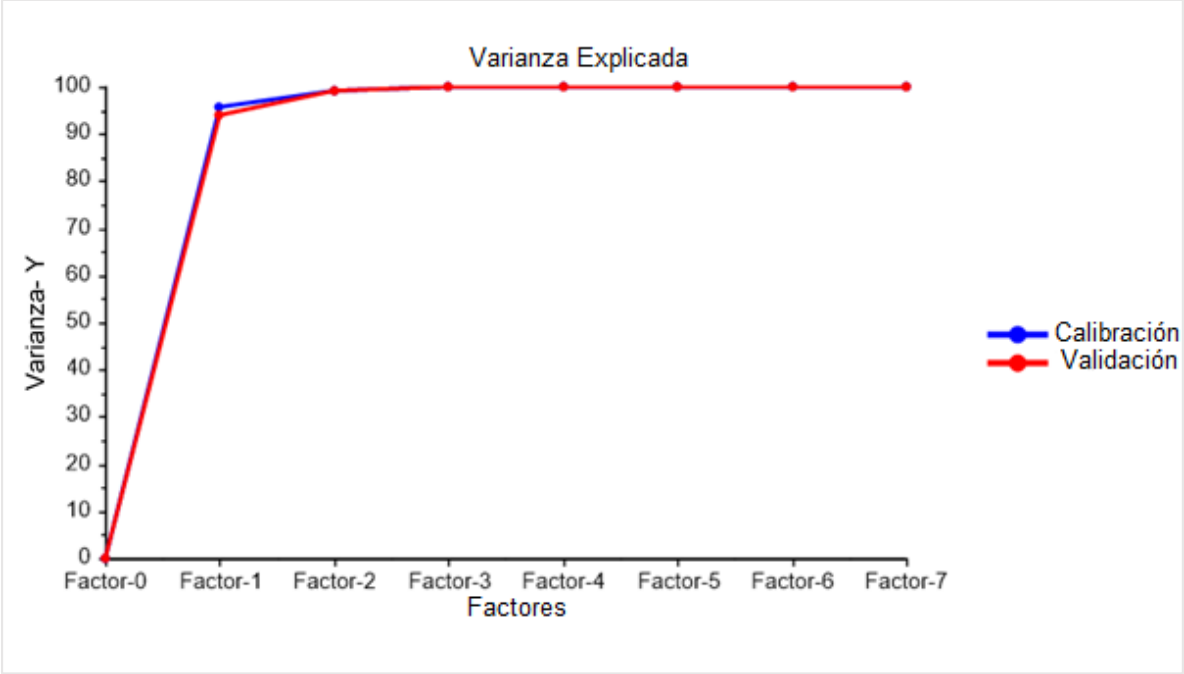


Figura 34. Gráfico de varianza en Y vs número de factores del PLS; Smoothing SGolay y corrección de línea base.

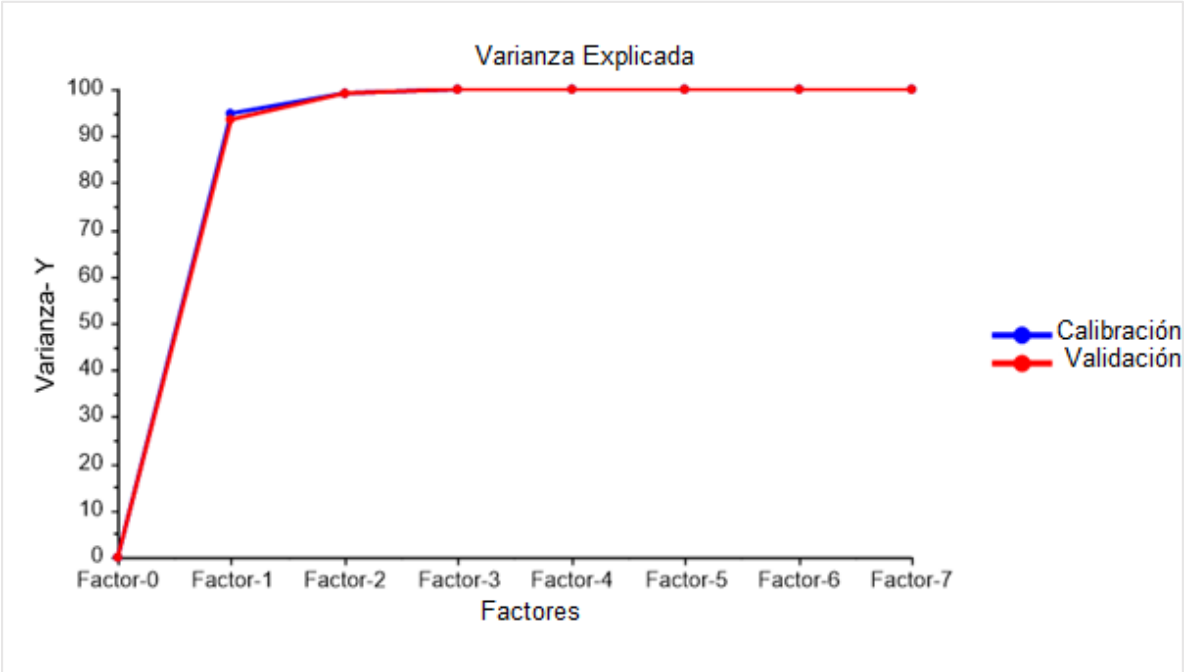


Figura 35. Gráfico de varianza en Y vs número de factores del PLS; algoritmo genético con Smoothing SGolay y corrección de línea base.

Si las varianzas explicadas (calibración y validación) difieren significativamente, existirían buenas razones para preguntarse si los datos de calibración son realmente representativos.

Por el contrario, si las dos curvas de varianza explicada están juntas los datos son representativos.

En la Figura 28 se observa que se necesitan 3 componentes principales para intentar tener una varianza explicada cercana a 100%, por ende, el modelo no se será lo suficientemente representativo, esta figura 28 hace referencia al modelo sin pretratamientos matemáticos y sin algoritmo genético.

Las Y-variables con una alta varianza explicada cercana al 100% (o pequeña varianza residual cercana a 0) son explicados con dos componentes principales como es el caso que se observa en las Figuras de la 29 a la 35 en donde los espectros han sido pre tratados y corridos por algoritmo genético, generando así modelos aún más representativos, lo cual significa que los datos de calibración se encuentran bien ajustados, por ende, el modelo describe bien los nuevos datos.

Las Figuras de la 36 a la 43 muestran los modelos de calibración multivariado desarrollado para la variable metanol con cuatro (4) factores para cada uno de los modelos diseñados.

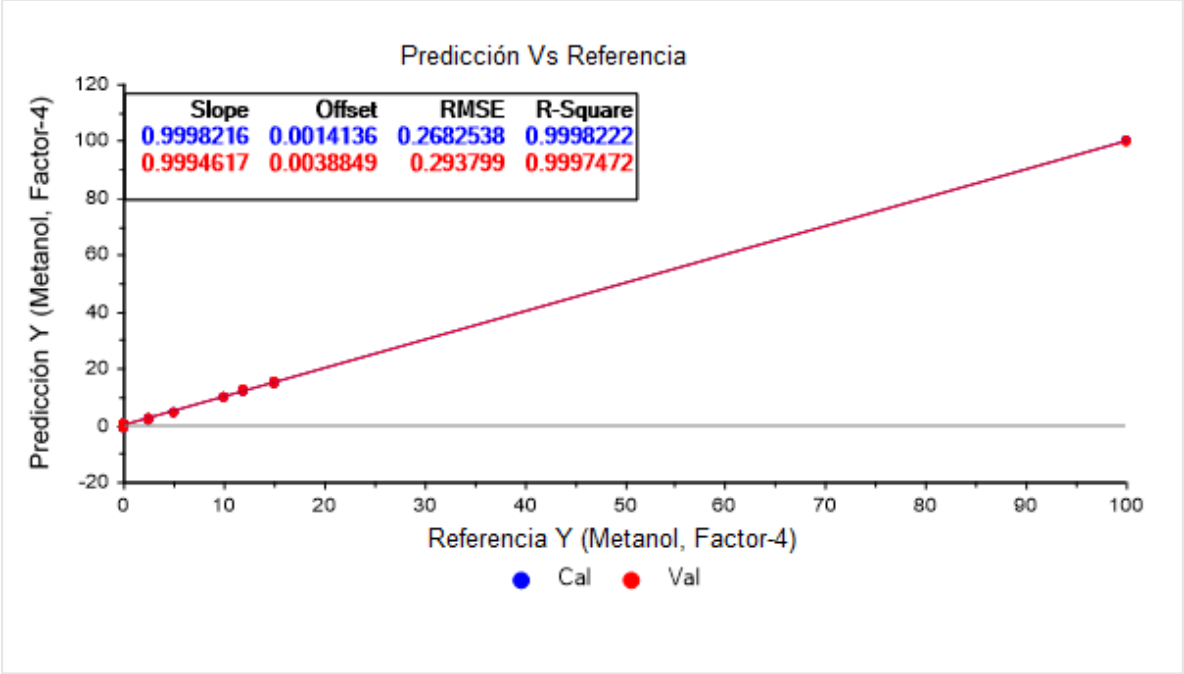


Figura 36. Gráfico de valores de Referencia vs Predicción para el modelo de calibración sin pretratamiento para metanol.

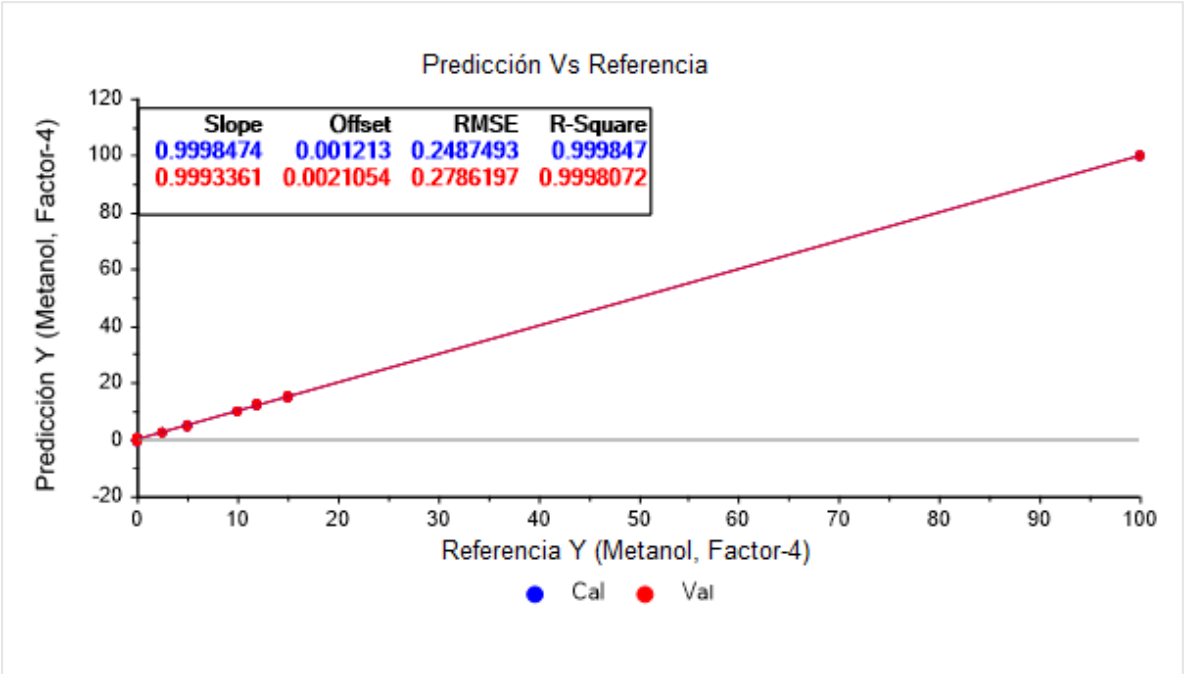


Figura 37. Gráfico de valores de Referencia vs Predicción para el modelo de calibración con algoritmo genético sin pretratamiento para metanol

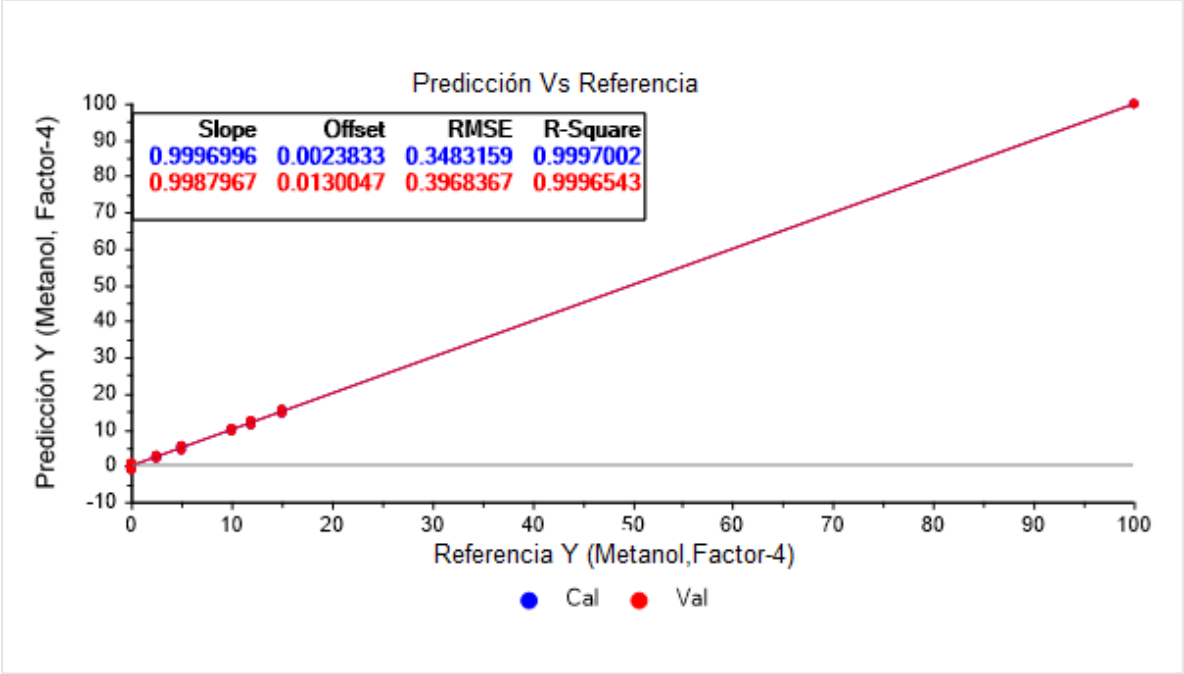


Figura 38. Gráfico de valores de Referencia vs Predicción para el modelo de calibración con Smoothing SGolay para metanol.

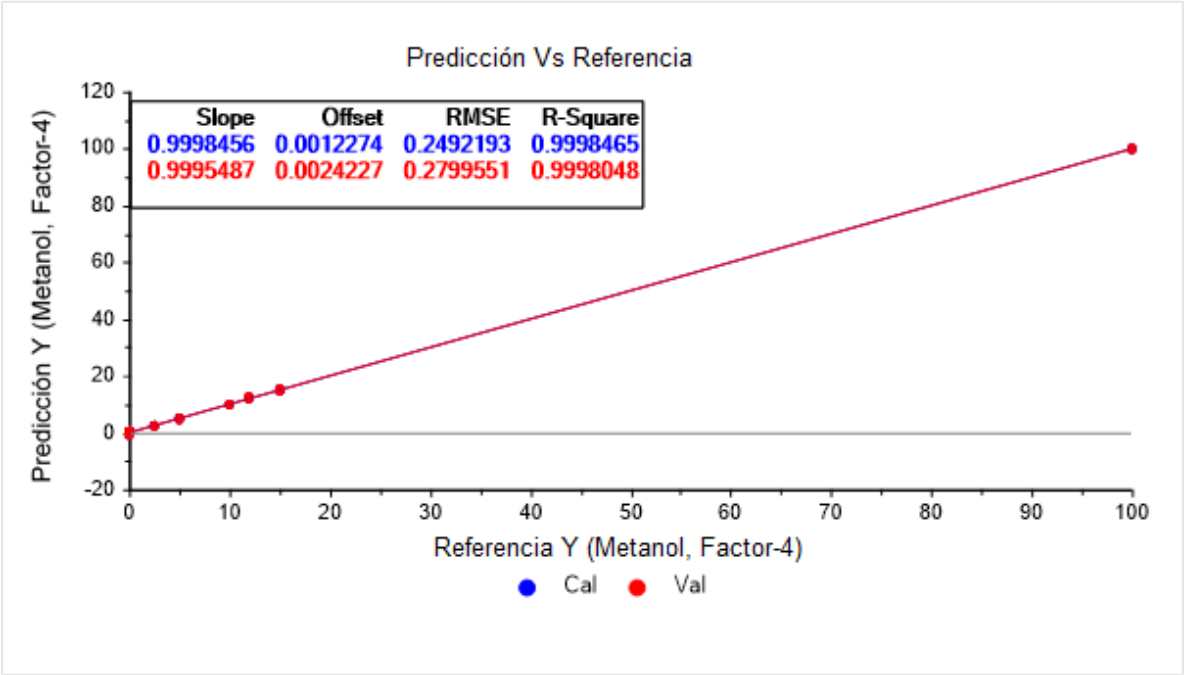


Figura 39. Gráfico de valores de Referencia vs Predicción para el modelo de calibración; algoritmo genético con Smoothing SGolay para metanol.

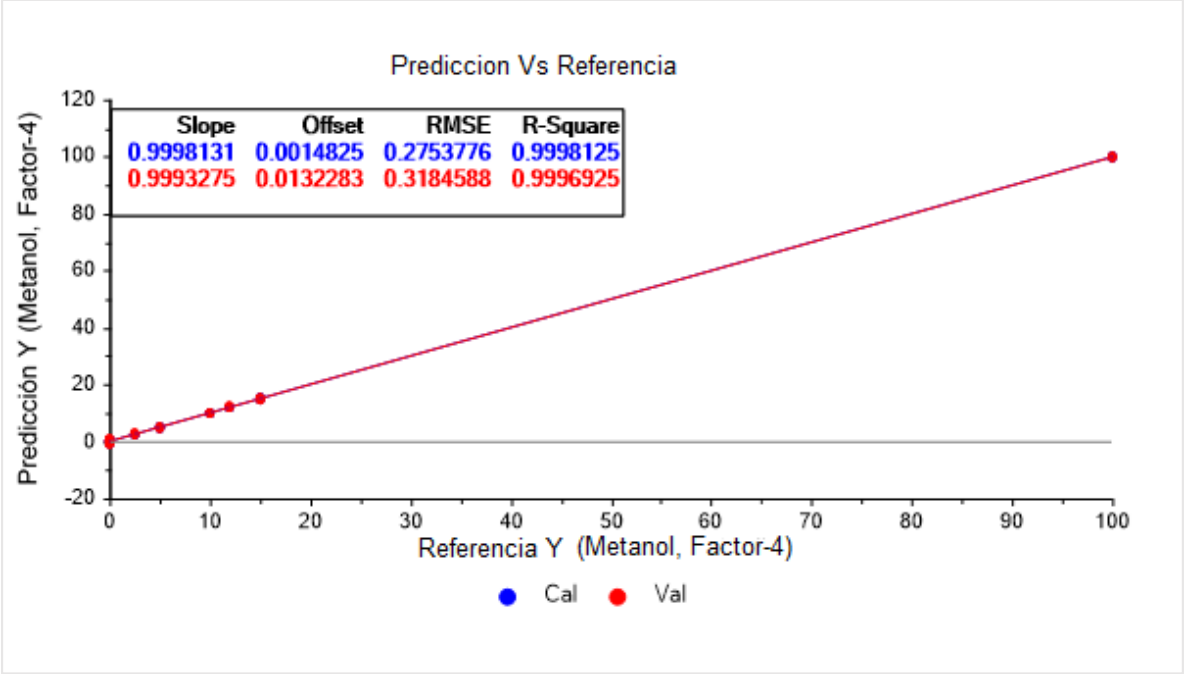


Figura 40. Gráfico de valores de Referencia vs Predicción para el modelo de calibración con corrección de línea base para metanol.

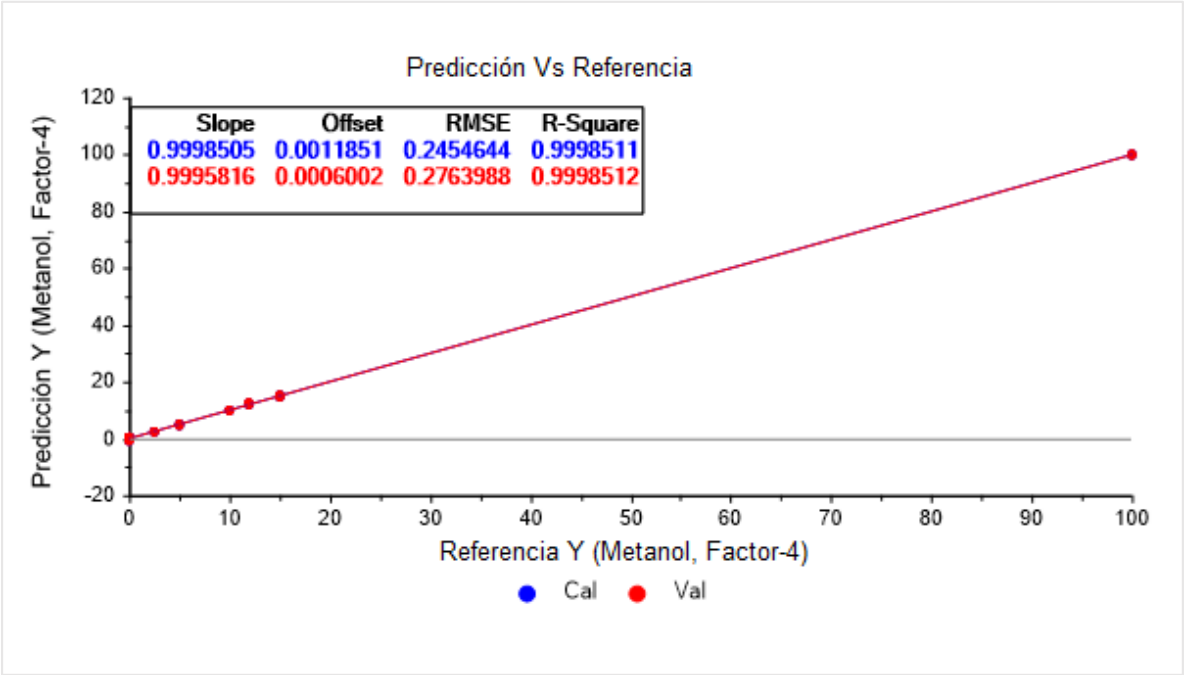


Figura 41. Gráfico de valores de Referencia vs Predicción para el modelo de calibración; algoritmo genético con corrección de línea base para metanol.

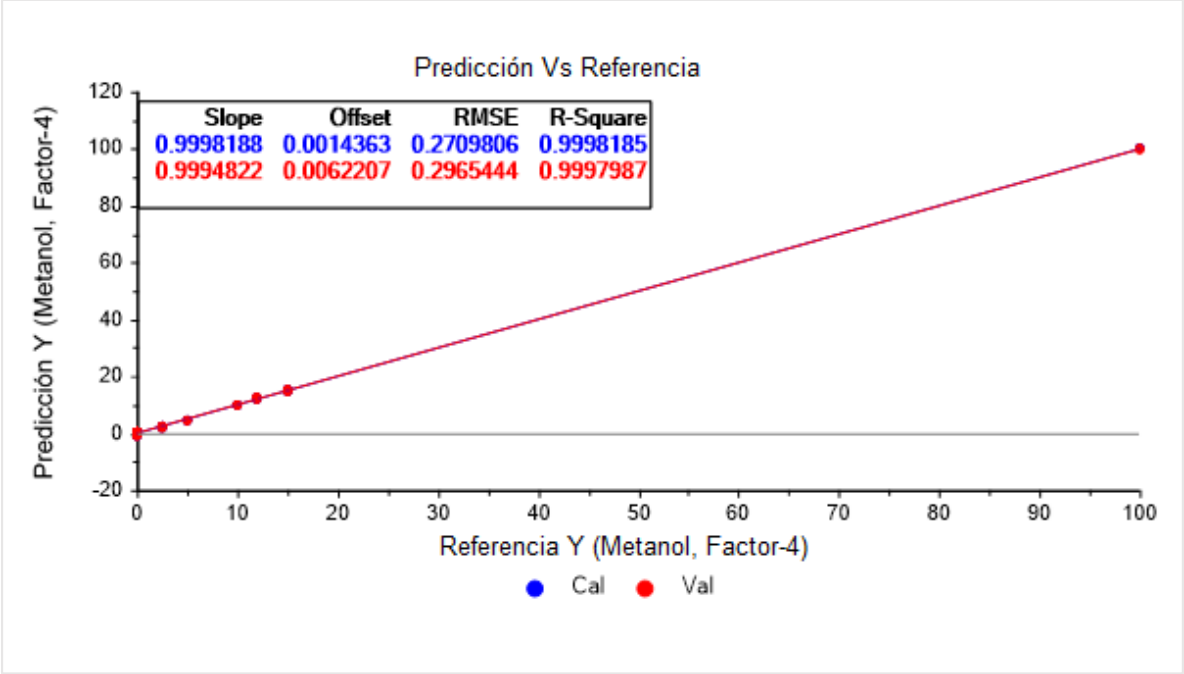


Figura 42. Gráfico de valores de Referencia vs Predicción para el modelo de calibración; Smoothing SGolay y corrección de línea base para metanol.

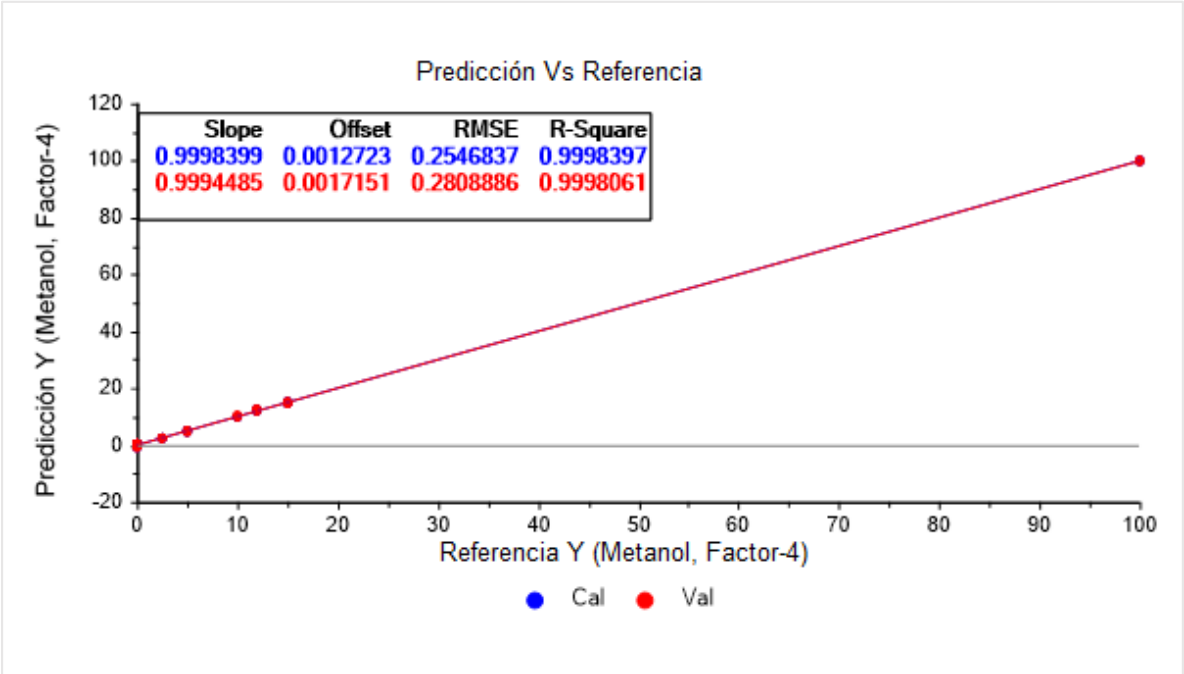


Figura 43. Gráfico de valores de Referencia vs Predicción para el modelo de calibración; algoritmo genético con Smoothing SGolay y corrección de línea base para metanol.

En las Figuras que van desde la 36 a la 43, La línea roja presentada en cada uno de las figuras de calibración multivariado hace referencia al conjunto de datos para la validación interna, en la que los datos escogidos para dicha validación son seleccionados al azar por el mismo software Unscrambler® X 10.4, mientras que la línea azul hace referencia al conjunto de datos de calibración en cada uno de los modelos multivariados desarrollados.

El R^2 varia de 0 a 1, en donde un valor de 0.9 se considera generalmente como bueno, esto varía dependiendo de la aplicación y la cantidad de muestras que se evalúan en el modelo.

Por otra parte, si la linealidad del modelo tanto de calibración como de validación interna se encuentran muy cerca como es el caso de las figuras 36 a la 43 en que los datos han sido pre tratados y aplicándoles algoritmo genético, el modelo será representativo y tendrá la capacidad para predecir muestras externas.

Las Tablas 4 y 5, muestran los resultados de validación interna con cada una de los pretratamientos mencionados anteriormente con y sin algoritmo genético para metanol.

Tabla 4. Resultados obtenidos modelos de calibración con algoritmo genético para metanol.

Pretratamientos	1F*		3F*		4F*	
	RMSE	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**
Sin tratamientos	13,81079	0,53862	0,48818	0,99941	0,24478	0,99984
Smoothing SGolay	13,81043	0,52859	0,35404	0,99969	0,24922	0,99984
Linea base	13,66268	0,52856	0,35332	0,99969	0,24546	0,99985
Smoothing SGolay-Baseline	13,66277	0,53861	0,48170	0,99942	0,25468	0,99983

*Número de factores de PLS empleados en los modelos

Tabla 5. Resultados obtenidos modelos de calibración para metanol.

Pretratamientos	1F*		3F*		4F*	
	RMSE	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**
Sin tratamientos	13,71083	0,53761	0,38105	0,99953	0,26825	0,99982
Smoothing SGolay	13,42637	0,55444	0,49373	0,99939	0,34831	0,99970
Linea base	13,67764	0,53536	0,43566	0,99966	0,27537	0,99981
Smoothing SGolay-Baseline	13,54336	0,54664	0,56081	0,99922	0,27098	0,99981

*Número de factores de PLS empleados en los modelos

El error medio cuadrático (RMSE) y el error estándar relativo (RSE) expresan la exactitud del modelo

El coeficiente de correlación de Pearson, pensado para variables cuantitativas, es un índice que mide el grado de covariación entre distintas variables relacionadas linealmente.

El coeficiente de correlación de Pearson es un índice de fácil ejecución e, igualmente de fácil interpretación, sus valores absolutos oscilan entre 0 y 1. En las calibraciones multivariables se evalúa la linealidad mediante la relación entre el valor de referencia y el valor estimado con el método desarrollado, mediante el cálculo del cuadrado del coeficiente de correlación de Pearson(R) (ecuación 7). cuanto más alto es el coeficiente, mejor será la correlación entre las medidas [43].

$$R = \frac{\sum_{i=1}^{n}(xi - \bar{x})(yi - \bar{y})}{\sqrt{[\sum_{i=1}^{n}(xi - \bar{x})^2]\sum_{i=1}^{n}(yi - \bar{y})^2]}} \qquad Ecuación\ 7$$

Cabe resaltar que cuando el cálculo del coeficiente de correlación se hace en la calibración del modelo, se hace referencia a este como R^2. En el caso de la validación, se hace referencia a este como Q^2.

En la Tabla 6 y 7 se comparan los valores de RMSE y R^2 aplicando los mismos tratamientos a cada una de las matrices de calibración con y sin algoritmo genético para metanol, en donde se puede observar claramente como el valor de RMSE va disminuyendo en cada una de las matrices de calibración con sus respectivos

tratamientos al aplicarle algoritmo genético y el R^2 se va haciendo más próximo a el valor esperado (1). El mismo patrón se puede observar para el etanol (Tabla 6 y 7). Se nota que el mejor modelo de calibración multivariado desarrollado para cada una de las variables de interés (metanol y etanol) se presenta con el modelo en donde se le aplico un pretratamiento de corrección de línea base y cuando esta serie de datos se someten al algoritmo genético, cabe resaltar que los coeficientes de correlación en cada uno de los modelos diseñados se encuentran superiores a 0.999 lo cual nos dice que las variables se encuentran relacionadas entre sí.

Tabla 6. Resultados obtenidos modelos de calibración con algoritmo genético para etanol.

Pretratamientos	1F*		3F*		4F*	
	RMSE	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**
Sin tratamientos	16,17342	0,30250	0,67276	0,99885	0,43409	0,99952
Smoothing SGolay	16,60970	0,30250	0,67188	0,99885	0,43641	0,99951
Baseline	16,83853	0,28315	0,42525	0,99954	0,41699	0,99956
Smoothing SGolay-Baseline	16,83897	0,28311	0,42533	0,99954	0,41746	0,99955

*Número de factores de PLS empleados en los modelos

Tabla 7. Resultados obtenidos modelos de calibración para etanol.

Pretratamientos	1F*		3F*		4F*	
	RMSE	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**
Sin tratamientos	13,71083	0,53536	0,74140	0,99861	0,44593	0,99949
Smoothing SGolay	16,88644	0,27906	0,74395	0,99860	0,43889	0,99951
Baseline	16,70844	0,29418	0,76708	0,99851	0,42304	0,99954
Smoothing SGolay-Baseline	16,80417	0,53861	0,92701	0,99782	0,45374	0,99947

*Número de factores de PLS empleados en los modelos

Las Tablas 8 y 9 hacen referencia a una comparación en los modelos de calibración desarrollados en este proyecto contra un modelo desarrollado en Brasil en el año 2007 y el método de referencia.

Tabla 8. Comparación resultados obtenidos en la calibración del modelo para metanol.

	SIN TRATAMIENTO		***CORRECCIÓN LÍNEA BASE***	
Modelo	RMSE	R^2	RMSE	R^2
Brasil	0.650	0.998	0.573	0.998
Colombia	0.268	0.999	0.245	0.999
Método referencia	0.5	0.995	-	-

Tabla 9. Comparación resultados obtenidos en la calibración del modelo para etanol.

	SIN TRATAMIENTO		***CORRECCIÓN LÍNEA BASE***	
Modelo	RMSE	R^2	RMSE	R^2
Brasil	0.787	0.993	0.814	0.993
Colombia	O.445	0.999	0.423	0.999
Método referencia	0.5	0.995	-	-

Teniendo en cuenta las comparaciones presentadas del error medio cuadrático (RMSE) como el coeficiente de correlación (R^2) en la Tabla 6 y 7, se observa que los modelos desarrollados en este estudio arrojaron mejores resultados en las predicciones de etanol y metanol con pretratamiento de corrección de línea base comparado con un estudio realizado en Brasil en el año 2007 [1] y con el método de referencia ASTM D5501-12 el cual es por cromatografía de gases [47].

Una vez que se selecciona un modelo y se comprueba su correlación en la representación de las muestras (Gasolina, Etanol y Metanol), es conveniente llevar a cabo una validación cruzada para así comprobar en qué medida el modelo se ajusta a otras posibles muestras pertenecientes a la misma población.

Teniendo en cuenta que el modelo de calibración multivariado escogido para la validación cruzada fue el que se obtuvo al aplicarle a la matriz original algoritmo genético y un pretratamiento de línea base debido a que el coeficiente de correlación (R^2) y al error medio cuadrático (RMSE) arrojados fueron los más óptimos.

La Tabla 10 y 11 muestra los datos de predicción Vs referencia para metanol y etanol presentes en cada una de las muestras preparadas.

La nomenclatura de las mezclas para llevar a cabo la validación cruzada es la siguiente, la M* hace referencia al componente metanol y la E* hace referencia al componente etanol, los números que acompañan las componentes hacen referencia a los porcentajes en cada una de las mezclas y las letras B* y C* hacen referencia a sus respectivas replicas.

Tabla 10. valores de predicción vs referencia para el metanol.

PREDICCIÓN (METANOL, FACTOR-4)	Y- PREDICCIÓN	Y- DESVIACIÓN	Y- REFERENCIA
M5	4.874486	0.2139	5
M5B	4.866261	0.2549	5
M5C	4.853464	0.2253	5
M5E7.5	4.984383	0.2831	5
M5E7.5B	5.007755	0.2175	5
M5E7.5C	5.012345	0.2345	5
M6E6	5.90582	01511	6
M6E6B	6.009655	0.1503	6
M6E6C	5.972354	0.1509	6
M7	6.936965	0.2442	7
M7B	6.945992	0.2218	7
M7C	6.954783	0.2386	7
M7.5E5	7.30288	0.1557	7.5
M7.5E5B	7.244907	0.1771	7.5
M7.5E5C	7.261423	0.1923	7.5

Tabla 11. valores de predicción vs referencia para el etanol.

PREDICCIÓN (ETANOL, FACTOR-4)	Y- PREDICCIÓN	Y- DESVIACIÓN	Y- REFERENCIA
M5	-0.099939384	0.3579	0
M5B	-0.1182861	0.3299	0
M5C	-0.1037551	0.3172	0
M5E7.5	7.472744	0.2864	7.5
M5E7.5B	7.398795	0.4133	7.5
M5E7.5C	7.526532	0.4209	7.5
M6E6	6.022476	0.2855	6
M6E6B	6.081194	0.2872	6
M6E6C	6.053182	0.2863	6
M7	-0.1463757	0.4641	0
M7B	-0.08337975	0.4214	0
M7C	-0.12887452	0.4025	0
M7.5E5	5.139952	0.3365	5
M7.5E5B	5.037129	0.2958	5
M7.5E5C	5.094325	0.3034	5

Las Figuras 44 y 45 muestran los resultados de la validación cruzada para los componentes (metanol y etanol) presentes en las muestras de gasolina preparadas. Teniendo en cuenta los parámetros obtenidos en la validación cruzada como lo son el coeficiente de correlación (R^2) superior a 0.999, un error medio cuadrático de predicción (RMSEP) y un error sistemático (Bias) relativamente bajo lo cual nos indica que tenemos una buena exactitud en la predicción de del etanol y metanol presentes en las muestras de gasolina previamente analizadas.

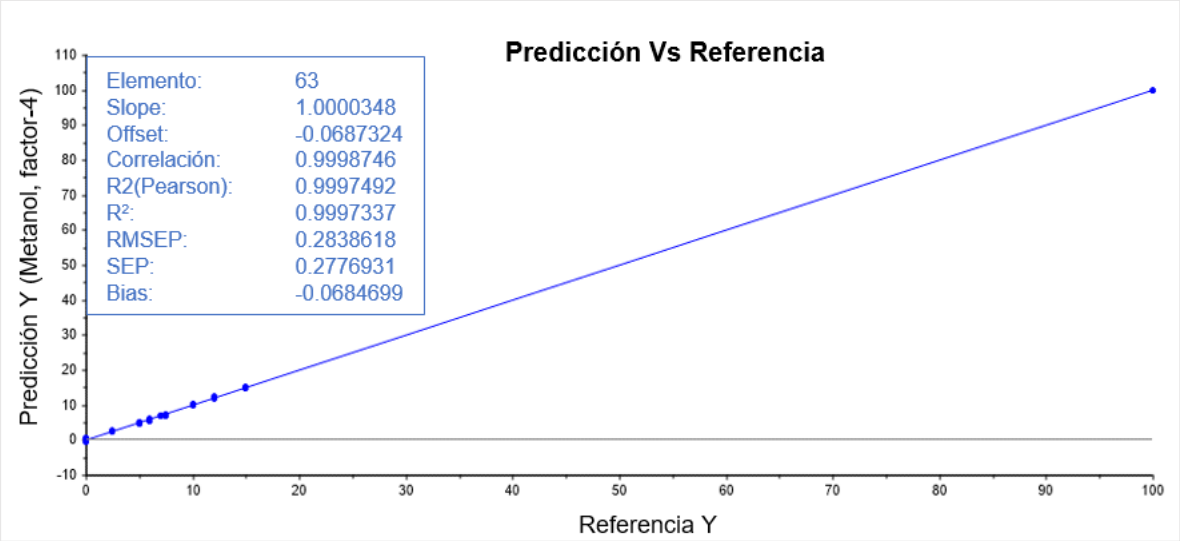


Figura 44. Modelo de validación cruzada para la determinación de metanol.

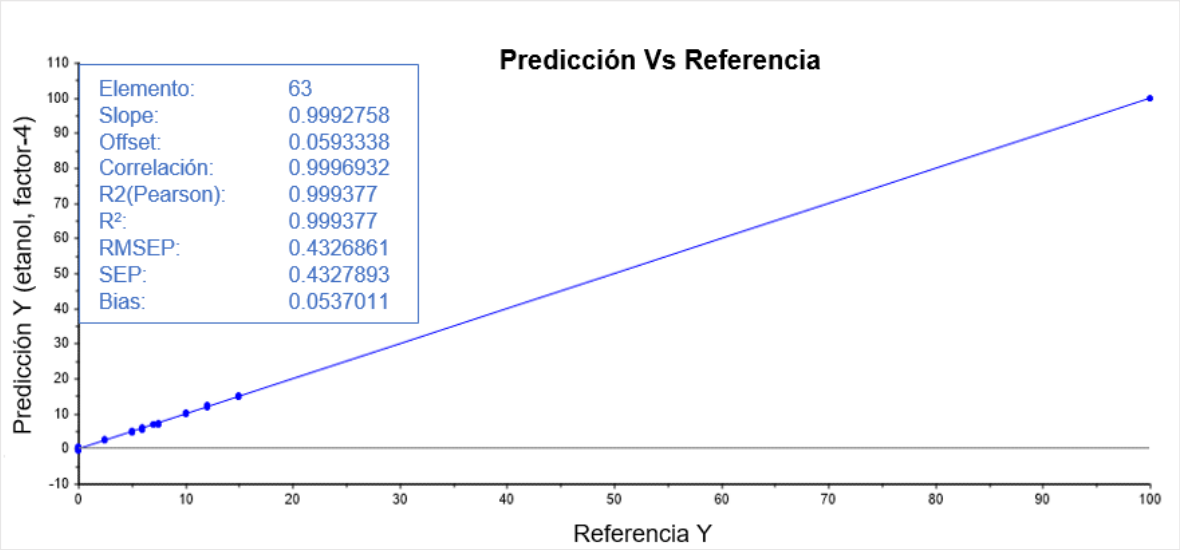


Figura 45. Modelo de validación cruzada para la determinación de etanol.

5. CONCLUSIONES

Los resultados obtenidos en este trabajo llevan a las siguientes conclusiones:

- La espectroscopia NIR junto con la calibración multivariada (PLS) pueden ser usado como método para la determinación simultánea de metanol y etanol en muestras de gasolina. Éste es método no destructivo y no es contaminante, siendo ésta su principal ventaja.

- Basándonos en el RMSE y el R^2 es posible observar que los modelos desarrollados por NIR presentan mejores resultados en comparación con el método de referencia (ASTM 5501-12).

- Existe una mejor predicción de metanol y etanol para aquellos modelos de calibración multivariados en los cuales se les aplicó algoritmo genético, teniendo en cuenta los pretratamientos.

- El mejor modelo para cuantificar y determinar tanto metanol como etanol en mezclas de gasolina fue aquel que se sometió a corrección de línea base, algoritmo genético y presentado con cuatro (4) factores.

6. RECOMENDACIONES

Validar muestras externas de distintas estaciones de servicio para verificar la correcta predicción del modelo desarrollado.

Es posible aumentar el porcentaje (v/v) tanto del metanol como del etanol para tener un mayor rango de identificación con respecto a las muestras.

Introducir muestras de producción o comercialización para hacer más robusto el modelo y mejorar su predicción.

7. REFERENCIAS BIBLIOGRAFICAS

[1] Fernandes H.L.; Raimundo I.M.; Pasquino C.; Rohwedder J.J. *Simultaneous determination of metanol and etanol in gasolina using NIR spectroscopy: Effect of gasoline composition.* Talanta 75 (2007). 804-810.

[2] Lutz O.; Bonn G.K.; Rode B.M.; Huck C.W.; *Reproducible quantification of etanol in gasolina via a customized mobile near-infrared spectrometer.* Anal. Chim. Acta 826 (2014). 61-68.

[3] Duarte V. *Especificaciones de la calidad del etanol carburante y del gasohol (mezcla de gasolina y etanol) y normas técnicas para la infraestructura*. Comisión económica para América Latina y el Caribe (CEPAL). 2016. 91.

[4] Cortes E.; González H.; Alvarez F. *Colombia en la era del alcohol carburante.* Ces 3 (2008). 120-132.

[5] Ministerio de Minas y Energía. Resolución No 40185. 2018.

[6] Balabin R.; Safieva R.; Lomakina E. *Gasoline classification using near infrared (NIR) spectroscopy data: Comparison of multivariate techniques.* Anal. Chim. Acta 671 (2010). 27-35.

[7] Corsetti S.; McGloin D.; Kiefer J. *Comparison of Raman and IR spectroscopy for quantitative analysis of gasoline/ethanol blends.* Fuels 166 (2016). 488-494.

[8] S. Naik, L. Meher, and D. Vidya, “Technical aspects of biodiesel production by transesterification-a review,” Renew. Sustain. Energy Rev., vol.10, pp. 248-268, 2006

[9] Santiago Arango A., Alina Torres F., *Economic Incidences of Ethanol as Biofuel in Colombia over the Sugar Cane Products.,* A System Dynamics Approach, Universidad Nacional de Colombia. 2008

[10] Helena García R., Laura Calderón E., *Evaluación de la política de biocombustibles en Colombia*, octubre 2012.

[11] Juan Eduardo Delgado., José Jorge Salgado., Ronaldo Pérez; *Perspectivas de los biocombustibles en Colombia*., universidad de Medellín, 2015.

[12] BIOCOMBUSTIBLES EN COLOMBIA: UN SECTOR EN CONSOLIDACIÓN., publicación del departamento de Biocombustibles Vol.4; 2012 Ecopetrol.

[13] PLAN INDICATIVO DE ABASTECIMIENTO DE COMBUSTIBLES LIQUIDOS, Unidad de Planeación Minero y Energético., Versión julio 2018.

[14] Ministerio de Minas y Energía., decreto 4892., 23 de diciembre 2011

[15] Referencia tomada de https://www.fedebiocombustibles.com/nota-web-id-487.htm

[16] ECOPETROL S.A., ICP. (2005) "*Efecto del etanol sobre las propiedades físico Químicas de las Gasolinas Colombianas y Desempeño en Motores y Vehículos* "Bogotá D.C.

[17] universitat Rovira i Virgili. Metodologías analíticas basadas en espectroscopía de infrarrojo y calibración multivalente. Aplicación a la industria petroquímica. Tesis doctoral, Tarragona, 2002.

[18] SKOOG, D.A.; Leary J.J., Holler F. James; *PRINCIPIOS DE ANÁLISIS INSTRUMENTAL*, 5° ed.; Ed. McGraw-Hill (1998), págs. 409-461.

[19] Brian G. Osborne., *Near-infrared Spectroscopy in Food Analysis*., the Encyclopedia of Analytical Chemistry in 2006 by John Wiley & Sons, Ltd.,2006, 1-14

[20] Davis, A. *An introduction to near infrared spectroscopy. NIRS* News Vol 16 Nº 7. http://www.nirpublications.com. (08 October 2006).

[21] C. Pasquini, "*Near Infrared Spectroscopy; Fundamentals, practical aspects and analytical applications,*" J. Brazilian Chem. Soc, vol.14. pp 196-219

[22] Murray, I. 1988. *Aspects of the interpretation of near infrared spectra*. Food Sci. Technol. Today. 2:135-139.

[23] Alomar, D. y R. Fuchslocher, 1998. *Fundamentos de la espectroscopía de reflectancia en el infrarrojo cercano (NIRS) como método de análisis de forrajes*. Agro Sur, 88-104

[24] Deaville, E. and P. Flinn, 2000. Near infrared (NIR) spectroscopy: an alternative approach for the estimation of forage quality and voluntary intake. p. 301-320.

[25] M. Gestal Pose., *Introducción a los algoritmos genéticos*, universidad de la coruña- España. Pp 2-3

[26] L. Davis (ed.) (1991). Handbook of Genetic Algorithms, Van Nostrand Reinhold, New York

[27] C. Reeves (1993). Modern Heuristic Techniques for Combinatorial Problems, Blackwell Scientific Publications

[28] Z. Michalewicz (1992). Genetic Algorithms + Data Structures = Evolution Programs, SpringerVerlag, Berlin Heidelberg.

[29] S. Fernández Henao, J. Soto Mejía., *Algoritmos genéticos aplicados en los sistemas de producción tipo pull*, universidad tecnológica de Pereira, agosto de 2011.

[30] *QUIMIOMÉTRIA, una disciplina útil para el análisis químico*, Universitat Rovira i Virgili, Tarragona España.

[31] Carlos Mongay Fernández, *Quimiométria*, universidad de valencia, 2005; pp 19-20

[32] Joan Ferré, *Calibración multivariable en análisis cuantitativo*, Universidad Rovira i Virgili, Tarragona España.

[33] A. Savitzky and M. Golay, *Smoothing and differentiation of data by simplified least squares procedures,* Anal. Chim., vol. 36, pp. 1627-1639. 1964

[34] G. Horlick*, Digital data handling of spectral utilizing Fourier transformations*, Anal. Chim., vol 44, pp. 943-947

[35] J. Workman, M. Koch, and D. Veltkamp*, process analytical chemistry*, Anal Chim., vol.77, pp 3789-3806. 2005

[36] J. Moros Portolés, *tratamiento numérico de los datos en el análisis cuantitativo por espectrometría vibracional*, universidad de valencia. 2007

[37] Ferreira, M., & Antunes, A. (1999). Quimiometria I: calibração multivariada, um tutorial. Quim. Nova, 724

[38] Martens, H., & Naes, T. (1996). Multivariate Calibration. John Wiley Sons.

[39] Sekulic, S., & Seasholtz, M. (1993). Nonlinear multivariate calibration methods in analytical chemistry. Anal. Chim, 835– 845.

[40] Anna Peguero Gutiérrez, *la espectroscopia NIR en la determinación de propiedades físicas y composición química de intermedios de producción y productos acabados*, universidad autónoma de Barcelona, 2010. Pp 66-69

[41] Santiago Macho Aparicio, *Metodologías Analíticas Basadas en Espectroscopia de Infrarrojo y Calibración Multivariante*. Aplicación a la Industria Petroquímica, Tesis Doctoral, Universitat Rovira i Virgili, Terragona, España, 2002.

[42] Multi-and Megavariate Data Análisis, Capitulo 3, pp 43-70

[43] J. Vega Vilca, J. Guzman; *regresión PLS y PCA como solución al problema de multicoliaidad en regresión múltiple*, 14 agosto 2010.

[44] Gemperline, P. (2006). Practical guide to chemometrics. New York: Taylor & Francis Group

[45] Gutiérrez J. *Aplicación de la Espectroscopia de Infrarrojo Cercano para la Determinación de Curvas de Destilación en Crudos de Carga en la Refinería de Barrancabermeja*., Universidad Nacional Abierta y a Distancia-UNAD. Bucaramanga. 2017

[46] Johmson R. *Applied multivariate statistical analysis.* Sexta edición. Prentice H. New Jersey. 2002. Cap. 1,9.

[47] Standard Test Method for Determination of Ethanol and Methanol Content in Fuels Containing Greater than 20% Ethanol by Gas Chromatography1