

CONSTRUCCIÓN DE MODELOS ESTADÍSTICOS PARA LA PREDICCIÓN DE
PROPIEDADES FISICOQUÍMICAS DE MUESTRAS DE CARBÓN COLOMBIANO
UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA INFRARROJA POR
TRANSFORMADA DE FOURIER (*FTIR-PAS*) Y REGRESIÓN DE MÍNIMOS
CUADRADOS PARCIALES (*PLS*)

YESID ROMÁN GÓMEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE INGENIERÍA FÍSICA
BUCARAMANGA
2017

CONSTRUCCIÓN DE MODELOS ESTADÍSTICOS PARA LA PREDICCIÓN DE
PROPIEDADES FISICOQUÍMICAS DE MUESTRAS DE CARBÓN COLOMBIANO
UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA INFRARROJA POR
TRANSFORMADA DE FOURIER (*FTIR-PAS*) Y REGRESIÓN DE MÍNIMOS
CUADRADOS PARCIALES (*PLS*)

YESID ROMÁN GÓMEZ

Trabajo de grado para optar al título de
Físico

Director

Dr. JÁDER ENRIQUE GUERRERO BERMÚDEZ

Codirector

M.Sc. RAFAEL CABANZO HERNÁNDEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE FÍSICA
BUCARAMANGA
2017

AGRADECIMIENTOS

Deseo expresar mi gratitud al profesor Jáder Guerrero por su dedicación, aporte de conocimiento, experiencia y paciencia al instruirme y guiarme durante todo el proceso de elaboración de este proyecto; al profesor Rafael Cabanzo por su apoyo, preocupación e interés en mi crecimiento académico, motivándome a mejorar y a Ximena Calderón por sus consejos que me facilitaron la elaboración de mi tesis de grado.

Finalmente deseo agradecer a mis padres, familiares y amigos por su apoyo moral y acompañamiento durante mis años de estudio.

CONTENIDO

	Página
1. PLANTEAMIENTO DEL PROBLEMA	15
2. OBJETIVOS	16
2.1. OBJETIVO GENERAL	16
2.2. OBJETIVOS ESPECÍFICOS	16
3. MARCO TEÓRICO	17
3.1. EL CARBÓN Y SU CLASIFICACIÓN	17
3.2. ESPECTROSCOPIA FOTOACÚSTICA INFRARROJA POR TRANSFORMADA DE FOURIER (<i>FTIR-PAS</i>)	19
3.3. QUIMIOMETRÍA Y MODELADO	21
4. EXPERIMENTAL	26
4.1. MUESTRAS Y MEDICIÓN DE PROPIEDADES	26
4.1.1 Determinación de humedad (%).	26
4.1.2 Prueba estándar para contenido de ceniza (%C).	27
4.1.3 Determinación del poder calorífico superior (PCS).	27
4.1.4 Determinación de la materia volátil (%MV).	29
4.1.5. Cálculo del Carbón fijo (%CF).	29
4.2. REGISTRO DE ESPECTRO <i>FTIR-PAS</i>	31
4.2.1 Parámetros instrumentales del espectrofotómetro.	32

4.2.2. Registro y pretratamiento de espectros.	33
5. ANÁLISIS Y RESULTADOS	35
5.1. CALCULO DE MODELOS POR <i>PLS-R</i>	35
5.1.1. Determinación de la complejidad y selección del modelo.	35
5.1.2. Diagnóstico y exactitud de los modelos.	39
6. CONCLUSIONES	45
REFERENCIAS BIBLIOGRÁFICAS	46
BIBLIOGRAFÍA	49

LISTA DE TABLAS

Tabla 1. Clasificación de carbones de acuerdo a normas de la <i>ASTM</i>	18
Tabla 2. Resultados de las propiedades fisicoquímicas de 28 muestras de carbón.	30
Tabla 3. Valores de la relación <i>SN</i> de espectros de muestras de carbón de la mina <i>Primavera</i> con distintos parámetros del equipo.	33
Tabla 4. Número de componentes principales, R^2 , <i>MSEP</i> y <i>RMSECal</i> para cada modelo seleccionado.	39
Tabla 5. Valores mínimos, máximo y promedio de las propiedades fisicoquímicas medidas.	40

LISTA DE FIGURAS

Figura 1. Esquema de un espectrofotómetro <i>FTIR</i>	20
Figura 2. Representación esquemática de la obtención de la señal fotoacústica..	20
Figura 3. Esquema de la bomba calorimétrica usada para la determinación del PCS	28
Figura 4. Celda fotoacústica acoplada al espectrofotómetro, vista exterior	31
Figura 5. Espectros <i>FTIR-PAS</i> tomados de las 28 muestras de carbón.	34
Figura 6. Espectros <i>FTIR-PAS</i> normalizados por z-score de las 28 muestras de carbón.	34
Figura 7. RMSEP de calibración contra RMSECal de predicción de las propiedades a) Ceniza y b) Carbón Fijo c) Humedad y d) Materia Volátil del e) Poder calorífico.	37
Figura 8. Cociente entre la dispersión de los errores (definida como la longitud del sector), respecto al error de sustitución contra el número de componentes para a) Ceniza y b) Carbón Fijo c) Humedad y d) Materia Volátil del e) Poder calorífico. .	38
Figura 9. Modelos seleccionados para a) Ceniza y b) Carbón Fijo c) Humedad y d) Materia Volátil del e) Poder calorífico.	41
Figura 10. Gráfica de predicciones por los modelos PLS contra los valores medidos para cada propiedad fisicoquímica.....	42
Figura 11. Residuales contra los valores de sus propiedades	43
Figura 12. Histograma de los residuales para todas las propiedades.....	44

RESUMEN

TÍTULO: CONSTRUCCIÓN DE MODELOS ESTADÍSTICOS PARA LA PREDICCIÓN DE PROPIEDADES FISICOQUÍMICAS DE MUESTRAS DE CARBÓN COLOMBIANO UTILIZANDO ESPECTROSCOPIA FOTOACÚSTICA INFRARROJA POR TRANSFORMADA DE FOURIER (FTIR-PAS) Y REGRESIÓN DE MÍNIMOS CUADRADOS PARCIALES (PLS)*

AUTORES: Yesid Román Gómez, Jáder Enrique Guerrero Bermúdez, Rafael Cabanzo Hernández**

PALABRAS CLAVES: Quimiometría, Carbón, Modelado, Espectroscopía infrarroja.

DESCRIPCIÓN: Colombia cuenta con grandes reservas de carbón, siendo este el segundo producto de exportación luego del crudo, además que cuenta con una participación en la producción mundial del 6.3%, situando a Colombia como el quinto exportador de carbón. El 95% de las reservas del país se ubican en los departamentos: Guajira, Cesar, Córdoba, Norte de Santander, Cundinamarca, Boyacá, Antioquia, Valle del Cauca y Cauca. Para determinar la calidad del carbón y clasificarlos, existen múltiples ensayos que sirven para conocer los diferentes contenidos elementales y propiedades fisicoquímicas, sin embargo, a pesar que se han desarrollado equipos automatizados para la realización de los procedimientos, estos requieren grandes cantidades de tiempo para la ejecución y pueden conllevar bastantes costos en su clasificación. De esta forma, aprovechando la alta sensibilidad a muestras características como el carbón, se utilizó la espectroscopía fotoacústica infrarroja con transformada de Fourier (*FTIR-PAS*) como una técnica útil y sencilla que, combinada con los métodos de análisis por regresión de mínimos cuadrados parciales (*PLS*) permite construir modelos de predicción a las respectivas propiedades fisicoquímicas de interés (humedad, ceniza, materia volátil, carbón fijo y poder calorífico), aliviando de este modo el problema de alto consumo de tiempo y minimizando los aportes de materiales contaminantes al medio ambiente.

* Proyecto de pregrado

** Facultad de Ciencias. Escuela de Física

Director: Dr. Jáder Enrique Guerrero Bermúdez. Codirector: M.Sc. Rafael Cabanzo Hernández

ABSTRACT

TITLE: STATISTICAL MODELS CONSTRUCTION FOR PHYSICOCHEMICAL PROPERTIES USING PHOTOACOUSTIC FOURIER TRANSFORM INFRARED SPECTROSCOPY (FTIR-PAS) AND PARTIAL LEAST SQUARE REGRESSION (PLS-R) *

AUTHORS: Yesid Román Gómez, Jáder Enrique Guerrero Bermúdez, Rafael Cabanzo Hernández **

KEYWORDS: Chemometrics, Coal, IR Spectroscopy, Analytical Chemistry.

DESCRIPTION: Colombia has large sources of coal and a share of 6.3% in the world's production. This get the country the fifth coal exporter. The coal is the second product of share after crude oil. The 95% of natural reserve are located between the departments of Guajira, Cesar, Córdoba, Norte de Santander, Cundinamarca, Boyacá, Antioquia, Valle del Cauca, and Cauca. The classification of coal is essential to decide its quality and use. Usually it is through determination of the elemental content and physicochemical properties. Currently, there are automated devices to measure their physicochemical properties. Nevertheless, the American Society for Testing and Materials (*ASTM*) establish procedures that are relatively time-consuming. The photoacoustic Fourier transform infrared spectroscopy (*FTIR-PAS*), has a characteristic sensibility for coal samples; and couple with partial least square regression (PLS-R) is an useful and simple technique to get physicochemical properties prediction models. Improving the difficulties of time and with minimum production of hazardous wastes to the environment. In this work, 28 samples of Colombian coal from several places were used to construction prediction models for moisture, ash, volatile matter, fixed coal and calorific value, with satisfactory results of root mean square errors (RMSE) and correlation coefficient (R^2).

* Degree Project

** Faculty of science. Physics School.

Adviser: Dr. Jáder Enrique Guerrero Bermúdez. Coadviser: M.Sc. Rafael Cabanzo Hernández

INTRODUCCIÓN

El carbón ocupa el segundo renglón en productos de exportación de Colombia, poniendo al país en el quinto lugar como exportador a nivel mundial y el primero en Latinoamérica, aportando de 1,5 % a 2 % en el PIB nacional con 85,5 millones de toneladas (Mt) en el 2015 [1] [2]. Según la Agencia Nacional Minera (ANM), *“El 90 % de la producción de carbón en Colombia es extraído en las zonas mineras del Cesar y La Guajira por empresas multinacionales como BHP-Billiton, Xtrata, Drummond, Glencore y Anglo-American, quienes desarrollan actividades de extracción a cielo abierto sobre las formaciones Cerrejón y Los Cuervos, ambas de edad Paleoceno-Eoceno. Estos carbones son de uso térmico, y en su mayor parte son exportados por los puertos de Santa Marta y Puerto Bolívar”* [3]. En esta perspectiva, la determinación de las propiedades fisicoquímicas del carbón ha desempeñado un papel significativo en la industria carbonífera, puesto que ha promovido la construcción de escalas de clasificación que permiten establecer precios en el mercado en forma diferencial. Estas propiedades determinan la calidad del carbón, estructura y hasta los posibles costos de transporte. Lo anterior ha motivado el desarrollo de distintos sistemas de clasificación, de acuerdo a la tecnología de la época. Así, en el siglo XIX preponderaba la apariencia visual, pero el incremento del uso industrial hizo necesaria la implementación de sistemas basados en las propiedades microscópicas y macroscópicas. En consecuencia se generaron varios sistemas como el propuesto por Seyler, el sistema internacional de clasificación, la clasificación NCB (National Coal Board) y el sistema Parr [4].

En efecto, el carbón es un combustible fósil resultado del proceso de carbogénesis o de transformación de materia orgánica, y es utilizado principalmente para combustión y coquizado, además de las múltiples aplicaciones tecnológicas entre las que se destaca: generación de electricidad, vapor de agua, en la industria

siderúrgica, cementera y en fabricación de coque. Inicialmente se considera un material orgánico (Por ejemplo, madera o algas) que es sometido a procesos leves de oxidación y putrefacción (Fase Bioquímica), posteriormente a lo largo de los años se realizan procesos de enterramiento que, dependiendo de sus condiciones como presión y temperatura, forzarán reacciones químicas y de esta forma altera la estructura (Fase metamórfica). La influencia de estas fases determinará los distintos contenidos de materia orgánica (Macerales) y la estructura del carbón, que son clasificados por la *ASTM* en primera parte como antracita, bituminoso, sub-bituminoso, lignito y turba, conocidos como los tipos de carbón [4].

Como se mencionó en el párrafo anterior, el carbón se utiliza en mayor parte para la combustión y coquizado; de esta forma las propiedades influyentes o significativas fueron inicialmente el contenido de carbono e hidrógeno, que obedecen leyes según su tipo estructural junto a propiedades fisicoquímicas como el poder calorífico y cenizas. Su estudio llevó al cálculo de volúmenes teóricos de aire y humos, al igual que porcentajes de oxígeno; de forma consecuente, Van Krevelen estudió las relaciones de H/C, O/C junto a propiedades como poder calorífico e índice de volátiles. Basándose en estos estudios, Parr establece que el poder calorífico e índice de volátiles son parámetros complementarios [4]. Esta característica básica es adoptada por la *American Society for Testing and Materials (ASTM)* para la clasificación del carbón. Así, la calidad comercial del carbón depende de propiedades tales como la capacidad calorífica, materia volátil, composición elemental, carbono fijo, cenizas y humedad.

Las técnicas espectroscópicas han jugado un papel importante en el estudio de materiales y particularmente en muestras de carbón, ya que permite estudiar y determinar unidades estructurales básicas, grupos funcionales, naturaleza del material mineral y contenido elemental. En general, son procedimientos que consumen poco tiempo, en comparación con las valoraciones convencionales. Por

otro lado, la espectroscopía con el concurso de la aplicación de la quimiometría ha potencializado la determinación de propiedades fisicoquímicas de muestras de carbón [5].

En este trabajo se calcularon modelos predictivos por el método de regresión de mínimos cuadrados parciales (*Partial Least Square Regression PLS-R*) con errores de predicción aceptables y coeficientes de linealidad superiores a 0.94 para las propiedades fisicoquímicas del carbón tales como ceniza, materia volátil, carbón fijo, humedad y poder calorífico; todo fue realizado utilizando espectroscopía infrarroja por transformada de Fourier en el rango infrarrojo medio por su significativo aporte de información respecto a la estructura del carbón. Además se utilizó una detección fotoacústica, por sus siglas en inglés *FTIR-PAS*, ya que presenta mayor sensibilidad en la toma de espectros para el carbón.

1. PLANTEAMIENTO DEL PROBLEMA

Conocer el grado de maduración o tipo del carbón puede realizarse mediante métodos químicos (conociendo el carbón fijo, cenizas, materia volátil, contenido de humedad y poder calorífico) como los propuestos por las normas *ASTM* o mediante las propiedades ópticas (espectroscopía *IR* y sus variantes) de los componentes del carbón, dado que el espectro infrarrojo contiene la información del contenido elemental, grupos funcionales, etc. El problema que nos concierne es la elaboración de modelos estadísticos, en el sentido de regresión de mínimos cuadrados parciales, para predecir las propiedades mencionadas arriba. Estos modelos se apoyan en una base de datos, constituida por 28 muestras distintas de carbón colombiano, cuyas propiedades han sido determinadas previamente por el grupo de investigación en espectroscopía atómica y molecular (LEAM) de la Universidad Industrial de Santander [6].

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Construir modelos de calibración multivariable para la predicción de propiedades fisicoquímicas de muestras de carbón colombiano, utilizando espectroscopía fotoacústica infrarroja por transformada de Fourier (*FTIR-PAS, Fourier Transform Infrared Photoacoustic spectroscopy*) y regresión de mínimos cuadrados parciales (*PLSR, Partial least square regression*).

2.2. OBJETIVOS ESPECÍFICOS

- Establecer los parámetros instrumentales (velocidad y número de barridos) que garanticen una adecuada resolución y razón señal a ruido para el registro de los espectros *FTIR-PAS* de las muestras de carbón.
- Formar una base de datos con los espectros *FTIR-PAS* de carbón proveniente de distintas regiones de Colombia y sus respectivas propiedades físico-químicas, las cuales han sido registradas en previos trabajos de grado.
- Construir modelos predictivos lineales, en el sentido de mínimos cuadrados parciales, para propiedades físico-químicas, entre otras: poder calorífico, porcentaje de ceniza, material volátil, carbón fijo.

3. MARCO TEÓRICO

3.1. EL CARBÓN Y SU CLASIFICACIÓN

El carbón es un agregado heterogéneo, estratificado y combustible que resulta de la acumulación y enterramiento de material vegetal y mineral. Este proceso recibe el nombre de *carbogénesis* y tiene un componente bioquímico y estructural; lo cual determina el tipo y rango de carbón. Para la formación del carbón es imprescindible que las sustancias vegetales no experimenten procesos agresivos de oxidación y putrefacción por microorganismos. [4] [6].

Ahora bien, las propiedades fisicoquímicas del carbón responden a variables como la temperatura y la presión, además de factores como: el contenido de nutrientes, oxígeno y el pH de la sustancia vegetal generadora. En particular, la temperatura a la cual estuvo sometido el material vegetal puede dar lugar a reacciones lentas necesarias para la formación y condensación de anillos polinucleares, que determinan la estructura (rango) del carbón. Teniendo en cuenta que la temperatura aumenta de 3 a 5 °C cada 100 metros de profundidad, la regla de *Hilt* indica que el rango aumenta regularmente con la profundidad de la capa terrestre, proporcionándose una relación profundidad-rango [4]. De este modo, las características fisicoquímicas del carbón, tales como la humedad, el poder calorífico, el porcentaje de materia volátil y contenido de material mineral no combustible (cenizas), guardan un estrecho compromiso con la génesis de una clase particular de carbón.

Las diferencias en las propiedades o características fisicoquímicas proporcionan un instrumento de clasificación para el carbón. En Colombia se ha adoptado la

clasificación de la *American Society of Testing and Materials (ASTM d-388)*. La Tabla 1 presenta rangos de las propiedades para diferentes tipos de carbón.

Tabla 1. Clasificación de carbones de acuerdo a normas de la *ASTM*

Tipo	Contenido de Humedad (%)	Material Volátil (%)	Carbono Fijo (%)	Poder Calorífico (MJ/kg)
Antracita	<15	1	86-98	>32.6
Bituminoso	15-20	32	45-86	24.5-32.6
Sub-bituminoso	20-30	50	35-45	18.2-24.5
Lignito y Turba	>30	96	25-35	9.3-18.2

A continuación se da una breve descripción de estas propiedades:

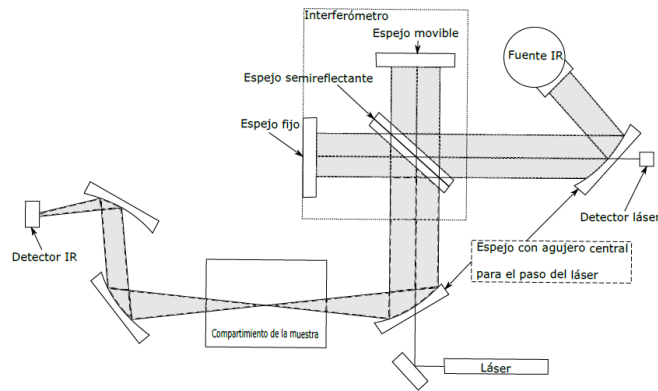
- El contenido de humedad (*ASTM D3173*) es el porcentaje de pérdida de peso al calentar la muestra entre 104°C y 110°C en la mufla, a esta temperatura se evita la eliminación de las aguas adheridas con un enlace químico fuerte. El valor porcentual de humedad determina su favorecimiento en la molienda, costo de transporte, la necesidad de auxiliares en su combustión (como ventiladores) y su nivel de aglomerado.
- El material volátil (*ASTM D3175*) se refiere a distintas cantidades de hidrogeno, óxidos de carbono, metano y otro hidrocarburos de bajo peso molecular que se desprenden en la fase inicial de combustión, usualmente se obtiene al calentar la muestra a 900 °C por 7 minutos. El conocimiento del material volátil indica la reactividad del carbón, su facilidad de ignición, estabilidad de la llama tras la ignición y la facilidad de su producción; por otra parte, una cantidad de materia volátil mayor al 30% puede causar problemas de seguridad. Además, en algunos tipos de carbón indica el rendimiento del coque producido y constituye un criterio de análisis en procesos de gasificación y licuefacción.

- El contenido de cenizas (*ASTM D3174*) es el residuo sólido tras la combustión completa de la materia orgánica y oxidación del material mineral, es una medida del contenido de sustancias minerales. Define la calidad del carbón ya que determina la cantidad de carbón incombustible y su rendimiento en la caldera, también indicando su poder calorífico (a mayor ceniza menor poder calorífico).
- El carbón fijo es la suma del contenido de humedad, material volátil y cenizas. Este factor define la combustión y carbonización e indica la cantidad de coque producible del carbón.
- El poder calorífico (*ASTM D5865*) se define como la cantidad de energía liberada cuando se incinera la unidad de masa en condiciones normalizadas de temperatura y presión. El poder calorífico superior usualmente se determina por la diferencia de temperaturas antes y después de la combustión en una bomba calorimétrica; realizando correcciones de humedad y contenido de los elementos *O, H, N, S* se obtiene el poder calorífico inferior.

3.2. ESPECTROSCOPIA FOTOACÚSTICA INFRARROJA POR TRANSFORMADA DE FOURIER (FTIR-PAS)

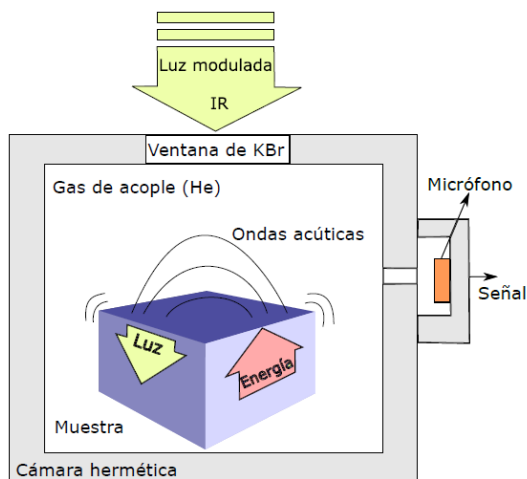
La espectroscopía infrarroja por transformada de Fourier es una técnica que emplea un interferómetro y una fuente de ancho de banda extenso en el rango infrarrojo. El patrón de interferencia (o radiación infrarroja modulada) se proyecta sobre la muestra y el espectro se obtiene mediante la transformación de Fourier discreta de la señal temporal que se registra mediante los detectores de infrarrojo (*IR*), ver Figura 1. El láser y el detector de radiación láser se utilizan específicamente para determinar la posición del espejo móvil.

Figura 1. Esquema de un espectrofotómetro *FTIR*



Ahora bien, si en lugar de registrar la radiación infrarroja dispersada o reflejada por la muestra, se miden las variaciones de presión, que se originan por los procesos de relajación no radiativos de la muestra, se obtiene un espectro a partir de la transformada de Fourier discreta de la señal fotoacústica (señal de presión). Naturalmente, en un montaje de espectroscopía fotoacústica, el sensor de radiación infrarroja se cambia por un sensor de presión. La técnica ha demostrado gran adaptabilidad y con resultados satisfactorios en sólidos, líquidos, geles, polvos y materiales biológicos. [7]

Figura 2. Representación esquemática de la obtención de la señal fotoacústica



La amplitud de la señal fotoacústica que se genera en la celda con atmósfera de gas inerte depende de la distribución de temperatura en el sistema gas-muestra-soporte, que atiende parámetros de la muestra, tales como: coeficiente de absorción óptica β , conductividad térmica k , calor específico c , densidad, difusividad térmica y la frecuencia de modulación de la radiación. Para explicar la generación de la onda acústica se dispone del modelo de Rosencwaig y Gersho (RG), en el cual la muestra transfiere energía térmica a las capas de gas más cercanas, que actúan como pistón; sin embargo el modelo ha sido extendido para considerar no sólo las oscilaciones del gas, sino las vibraciones de la superficie de la muestra. Cabe mencionar que la información obtenida con este método, por lo tanto los espectros, es proveniente de la energía absorbida, de ahí que es bastante similar a otros métodos comunes espectroscópicos de absorbancia.

3.3. QUIMIOMETRÍA Y MODELADO

Se entiende por *quimiometría* el conjunto de procedimientos y técnicas tomadas de la estadística, el diseño de experimentos, la minería de datos, el tratamiento de señales y estrategias de inteligencia artificial (entre otras: redes neuronales artificiales, lógica difusa, máquinas de soporte vectorial y algoritmos genéticos), aplicados a la química, que mediante el establecimiento de relaciones entre un conjunto de factores y respuestas, pretende clasificar, reconocer patrones, optimizar y construir modelos de calibración.

En el conjunto de herramientas estadísticas para la construcción de modelos predictivos se encuentran las técnicas de regresión. En este trabajo se enfatiza el análisis por mínimos cuadrados parciales (*PLS, Partial Least Square*), el cual permite establecer correlaciones entre múltiples variables de entrada y múltiples propiedades o entre múltiples variables de entrada y una única propiedad. Particularmente, las variables de entrada o predictores son el contenido espectral

en el infrarrojo medio ($400 \text{ cm}^{-1} - 4000 \text{ cm}^{-1}$) de muestras de carbón, provenientes de diferentes regiones de Colombia y la respuesta del modelo es una propiedad fisicoquímica de la muestra, ya sea el poder calorífico, el contenido de ceniza, carbón fijo o material volátil.

La conjugación de técnicas espectrales, de diferente naturaleza, y la quimiometría, presenta una tendencia en ascenso porque alivia procedimientos analíticos que consumen gran cantidad de tiempo e insumos. Además, es relativamente fácil la adaptación para el análisis remoto y simultáneo de la información.

En la elaboración de un modelo quimiométrico por *PLS* se recomienda, dependiendo de los datos registrados, un pretratamiento que comprende: el suavizado, el cual incrementa la razón señal ruido, el centrado de datos que disminuye tendencias en los datos, la corrección de línea base y la normalización que propende por una escala más uniforme y aportan un margen de independencia de las condiciones particulares del registro de los espectros.

En un modelo apoyado en *PLS*, se obtiene un grupo de variables latentes usando dos bloques de datos: las variables independientes (predictores), en este caso la matriz $\mathbf{X}_{m \times n}$ formada por m espectros de infrarrojo de n número de ondas (la dimensión de un espectro es $1 \times n$); y las variables dependientes $\mathbf{y}_{m \times 1}$, formado por los valores de referencia de las respectivas propiedades fisicoquímicas de las m muestras de carbón. Los valores de referencia se han obtenido, previamente, aplicando los procedimientos establecidos por las normas *ASTM*. Estas variables latentes se obtienen mediante algoritmos iterativos que maximizan la covarianza entre ambos bloques de datos.

De manera general, el modelo relaciona las respuestas predichas $\hat{\mathbf{y}}_{m \times 1}$ y los predictores (variables de entrada) mediante la expresión (1).

$$\hat{\mathbf{y}}_{m \times 1} = \mathbf{X}_{m \times n} \mathbf{b}_{n \times 1} \quad (1)$$

Donde $\mathbf{b}_{n \times 1}$ es el vector de coeficientes del modelo. El vector $\mathbf{b}_{n \times 1}$, se obtiene por algoritmos iterativos bien conocidos, como el *NIPALS (Nonlinear Iterative Partial Least Square)* o el *SIMPLS (Statistically Inspired Modification of Partial Least Square)*.

A manera de ejemplo, se describe a continuación el algoritmo *NIPALS-I*, que básicamente obtiene el mismo resultado que el algoritmo *SIMPLS*, cuando se trata de una única variable dependiente.

Inicialmente, la matriz de espectros $\mathbf{X}_{m \times n}$ se centra respecto al espectro promedio, esto es:

$$\mathbf{X} = \mathbf{X}_{\text{inicial}} - \mathbf{X}_{\text{promedio}} \quad (2)$$

La idea es iterar sobre un número de componentes principales, de manera que satisfaga criterios de error y ajuste. El bucle es el siguiente:

1. Se asigna al vector $\mathbf{u}_{m \times 1}$ el arreglo $\mathbf{y}_{m \times 1}$, que son los valores de referencia de la propiedad que se quiere predecir.

$$\mathbf{u}_{m \times 1} = \mathbf{y}_{m \times 1} \quad (3)$$

2. Calculamos $\mathbf{w}_{1 \times n}^T$, que conforma la matriz de pesos $\mathbf{W}_{n \times d}$, donde d es el número de componentes principales.

$$\mathbf{w}_{1 \times n}^T = \frac{\mathbf{u}_{1 \times m}^T \mathbf{X}_{m \times n}}{\mathbf{u}_{1 \times m}^T \mathbf{u}_{m \times 1}} \quad (4)$$

El vector peso calculado arriba, se divide por su norma.

3. Posteriormente se estima el vector $\mathbf{t}_{m \times 1}$ que compone la llamada matriz de *score* de $\mathbf{X}_{m \times n}$

$$\mathbf{t}_{m \times 1} = \mathbf{X}_{m \times n} \mathbf{w}_{n \times 1} \quad (5)$$

4. Se calculan los *loadings* de $\mathbf{y}_{m \times 1}$:

$$q = \frac{\mathbf{t}_{1 \times m}^T \mathbf{y}_{m \times 1}}{\mathbf{t}_{1 \times m}^T \mathbf{t}_{m \times 1}} \quad (6)$$

Nótese que para una única variable dependiente ($m = 1$), q es un escalar.

5. Se genera un nuevo vector $\mathbf{u}_{m \times 1}$ dividiendo el vector $\mathbf{y}_{m \times 1}$ por q .

6. Se calculan *loadings* de $\mathbf{X}_{m \times n}$:

$$\mathbf{p}_{1 \times n}^T = \frac{\mathbf{t}_{1 \times m}^T \mathbf{X}_{m \times n}}{\mathbf{t}_{1 \times m}^T \mathbf{t}_{m \times 1}} \quad (7)$$

7. Se obtienen los residuales de $\mathbf{X}_{m \times n}$

$$\mathbf{E}_{m \times n} = \mathbf{X}_{m \times n} - \mathbf{T}_{m \times d} \mathbf{P}_{d \times n}^T \quad (8)$$

Donde $\mathbf{T}_{m \times d}$ y $\mathbf{P}_{d \times n}^T$ corresponden a las matrices formadas por los vectores $\mathbf{t}_{m \times 1}$ y $\mathbf{p}_{1 \times n}^T$, respectivamente. Igualmente se calculan los residuales (vector $\mathbf{F}_{m \times 1}$) de $\mathbf{y}_{m \times 1}$

$$\mathbf{F}_{m \times 1} = \mathbf{y}_{m \times 1} - \mathbf{T}_{m \times d} \mathbf{q}_{d \times 1} \quad (9)$$

8. Los coeficientes del modelo se determinan mediante:

$$\mathbf{b}_{n \times 1} = \mathbf{W}_{n \times d} (\mathbf{P}_{d \times n}^T \mathbf{W}_{n \times d})^{-1} \mathbf{q}_{d \times 1} \quad (10)$$

9. Finalmente se valida el modelo para un número de componentes determinado, de acuerdo a algún criterio de selección, calculando y analizando las respuestas predichas $\hat{\mathbf{y}}_{m \times 1}$. Entre las formas de validación cruzada se encuentra el método *Montecarlo*, que consiste en una selección aleatoria de un porcentaje de las muestras para la calibración y otro para la predicción. La literatura recomienda que el número de modelos que se deben generar para seleccionar un “*óptimo*” desempeño debe ser mayor al orden del cuadrado del número de muestras.

Además, cuando se dispone de pocas muestras, el porcentaje de las usadas para validar deber ser por lo menos del 40%.

Por otro lado, las herramientas de diagnóstico del desempeño de los modelos de predicción, más ampliamente usadas, comprenden el análisis de residuales, el coeficiente de determinación R^2 y el comportamiento del error cuadrático medio de predicción (o la raíz cuadrada de este error), ecuación (11). Donde los $F_{i \times 1}$ son los residuales, ecuación (9) y m^* es el número de muestras de predicción o calibración.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m^*} (F_{i \times 1})^2}{m^*}} \quad (11)$$

4. EXPERIMENTAL

4.1. MUESTRAS Y MEDICIÓN DE PROPIEDADES

En el desarrollo de este proyecto se utilizaron 28 muestras de carbón suministradas por el laboratorio de geoquímica del petróleo de la Universidad Industrial de Santander, provenientes de distintas regiones de Colombia.

Cada muestra se pulverizó hasta pasar por un tamiz de 250 μm (Tamiz No.60), ajustándose al tratamiento sugerido en la norma *ASTM D346* o método D2013. En trabajos previos, se midieron las propiedades fisicoquímicas de interés: humedad (*ASTM D3173*), poder calórico (*ASTM D5865*), cenizas (*ASTM D3174*), material volátil (*ASTM D3175*) y carbono fijo (*ASTM D3174*) [6] [8] [9] [10] [11]. Parte de cada una de las muestras fueron destinadas al análisis espectral *FTIR-PAS*.

A continuación, se hace una breve descripción del procedimiento para medir las propiedades fisicoquímicas enumeradas en el párrafo anterior.

4.1.1 Determinación de humedad (%). Para la medición del contenido o porcentaje de humedad, se determina la masa s_1 del crisol totalmente seco y su respectiva tapa; luego se agrega 1 *g* de cada muestra obteniéndose una masa s_2 (es decir, s_2 es s_1 más un gramo de carbón). La muestra en su crisol destapado se somete a 110 °C por aproximadamente una hora en la mufla. Luego se retira de la fuente térmica y el crisol tapado se inserta en un desecador por treinta minutos hasta alcanzar la temperatura ambiente. Finalmente se determina la masa s_3 de cada contenedor (tapado) con la muestra después del proceso de secado. La expresión (12) permite el cálculo del contenido de humedad.

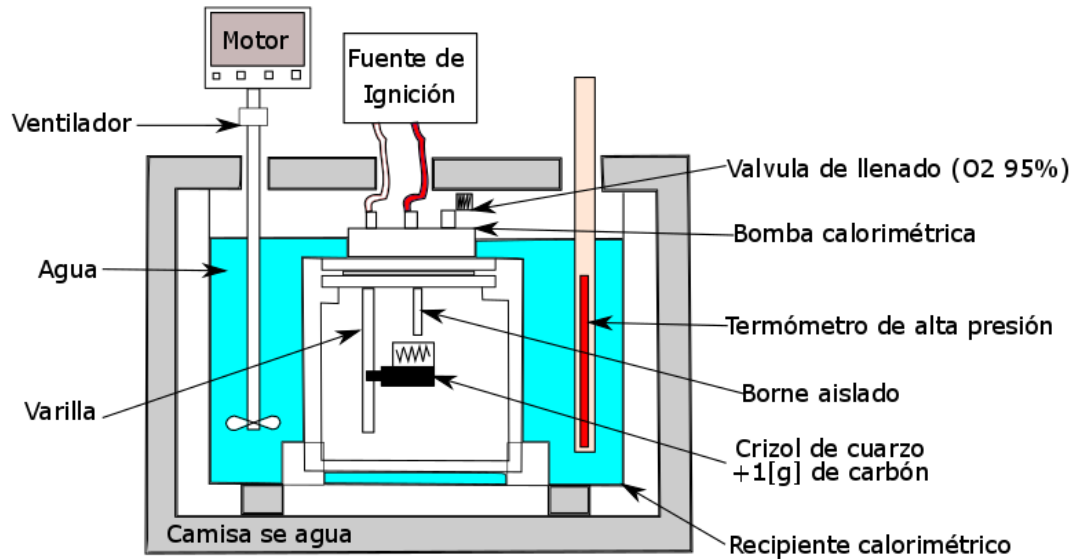
$$\%H = \frac{s_2 - s_3}{s_2 - s_1} * 100 \quad (12)$$

4.1.2 Prueba estándar para contenido de ceniza (%C). Los crisoles sin tapa, con la muestra desprovista de humedad (masa s_3), se ponen en la mufla; lentamente se sube la temperatura hasta alcanzar los 700 °C y se mantienen en combustión por dos horas. Posteriormente se retiran y se dejan enfriar sobre una placa de cerámica para resistir el choque térmico. Se trasladan al desecador por 45 minutos más, para luego determinar su masa s_4 . Así, con la ecuación (13) se calcula porcentaje de ceniza.

$$\%C = \frac{s_4 - s_1}{s_3 - s_1} * 100 \quad (13)$$

4.1.3 Determinación del poder calorífico superior (PCS). Para medir el poder calorífico superior, una muestra de carbón, de masa m , en un crisol de cuarzo, se cubre con algodón y se coloca en una bomba calorimétrica. La muestra se somete a combustión en presencia de oxígeno a 30 atmósferas. La ignición se produce a través de un filamento de Ni-Cr, de 10 cm. Antes de la ignición, se proporcionan 2.5 L de agua a la bomba. Durante la combustión se mantiene en agitación constante el agua y se registra periódicamente su temperatura hasta alcanzar el equilibrio térmico por cinco minutos.

Figura 3. Esquema de la bomba calorimétrica usada para la determinación del PCS



Los gases residuales se liberan y la cámara de ignición se lava con metil-naranja hasta no observar más reacción ácida. El poder calorífico se calcula mediante la expresión (14).

$$PCS = \frac{\Delta T * C - E_1 - E_2 - E_3}{m} \quad (14)$$

En la expresión (14), ΔT es el incremento desde la temperatura inicial del agua hasta alcanzar el equilibrio térmico. C , es la capacidad térmica del calorímetro, en particular $5088.36 \left[\frac{BTU * g}{lb * ^\circ C} \right]$, E_1 corresponde al calor de combustión del algodón ($7524 * \text{peso del algodón usado en } lb [BTU * g/lb]$), E_2 es el calor de combustión del filamento fusible Ni-Cr y E_3 es el calor de formación del ácido sulfúrico.

4.1.4 Determinación de la materia volátil (%MV). Los crisoles vacíos y con tapa se sometieron por un periodo de siete minutos a una temperatura entre 910 °C y 920 °C; se dejaron reposar por 45 minutos en el desecador para luego ser pesados. Esto se hace con fin de eliminar materia volátil adherida de experimentos anteriores; posteriormente se agrega una muestra de 1 g de carbón y se realiza el mismo procedimiento. Finalmente, el porcentaje de materia volátil en las muestras se calcula de acuerdo a la siguiente expresión.

$$\%MV = \frac{s_2 - s_3}{s_2 - s_1} * 100 - \%H \quad (15)$$

Donde s_1 corresponde a la masa en gramos del crisol seco con tapa y sin carbón, s_2 es la masa en gramos del crisol con 1 g de carbón y tapa, antes del secado, y s_3 es la masa del crisol con 1 g de carbón y tapa después del secado.

4.1.5. Cálculo del Carbón fijo (%CF). Por último, el carbón fijo se calcula con la diferencia a 100 de los porcentajes de contenido de humedad, ceniza y materia volátil.

$$\%CF = 100 - \%H - \%C - \%MV \quad (16)$$

Ajustados al protocolo descrito, en la Tabla 2, se resumen las valoraciones de las propiedades fisicoquímicas (Tomado de [6]).

Tabla 2. Resultados de las propiedades fisicoquímicas de 28 muestras de carbón.

Nombre de la mina	%H Wt%	%C Wt%	%MV Wt%	%CF Wt%	PCS BTU/lb
Hatillo	2,95	9,10	40,49	47,47	11495,20
Barranca	1,49	3,64	40,09	54,79	13501,70
Cúcuta	0,28	2,29	36,45	60,98	13601,21
La Esperanza Manto-Quintana 1	0,64	16,85	31,07	51,44	12876,01
<i>Drummond</i> de Loma Manto-Malba 8	5,63	5,04	38,81	50,52	12253,12
Primavera 14-Grande	2,95	6,77	38,24	52,04	12617,62
El Proceso Manto-8	0,35	4,18	35,41	60,06	13496,93
Santa Marta	1,62	5,02	39,19	54,17	12830,61
Yerbabuena Manto-10 Interior	5,63	4,98	41,21	48,17	12503,74
Zarzal	5,61	15,73	34,18	44,48	12149,21
El Zarzal Manto 1,40	6,03	10,27	32,98	50,72	12586,31
Kamelia Manto-20 (Zulia)	7,11	10,49	32,37	50,02	12171,13
Beta Grande Manto-6	0,59	5,42	34,54	59,45	13544,13
Cuaron Manto-Vera grande	0,68	7,51	34,66	57,16	13167,06
Diamante Manto Quintana TPH-3	0,85	1,74	36,82	60,59	13906,79
El Cerezo	1,52	3,00	32,28	63,20	14382,87
El Tesoro Manto-Meta chica	2,65	2,24	32,05	63,05	14521,55
La Esperanza la Rebelion Manto-5 Quintana y Ciscuda	0,88	3,48	35,44	60,20	13869,06
La Esperanza Procina Manto- Ciscuda 2	0,27	4,08	35,74	59,91	13934,31
La Esperanza TPH-3 Manto- Ciscudo	0,28	5,29	36,04	58,39	13611,50
Fortaleza	1,95	4,79	36,88	56,38	13689,47
Jamundi la Ferreira Manto-8	0,88	5,31	44,29	49,52	13093,04
La Lega Manto-2	0,59	9,21	31,57	58,63	13685,58
San Antonio Manto-1	1,72	4,24	30,93	63,12	14507,46
San Antonio Manto-2	6,84	5,09	28,64	59,43	14231,57
Mina 2	6,47	0,66	40,85	52,02	13059,96
Muestra X LEAM	4,84	1,43	41,78	51,95	12859,96
Yerbabuena Manto-40	3,43	6,00	41,04	49,53	12759,96

4.2. REGISTRO DE ESPECTRO *FTIR-PAS*

Los espectros de *FT-IR PAS* se registraron mediante una cámara fotoacústica (*MTEC* modelo 300) acoplada a un espectrofotómetro *Thermo Scientific Nicolet iS50 FT-IR*, ver Figura 4. La cámara posee un compartimiento para colocar la muestra, ver Figura 2. En la celda se coloca la muestra y se instaura una atmósfera de helio, antes de ser cerrada. Para un registro correcto de los espectros *FTIR-PAS*, es recomendable determinar parámetros instrumentales, tales como: ganancia, flujo de Helio, velocidad y número de barridos o escaneos.

Figura 4. Celda fotoacústica acoplada al espectrofotómetro, vista exterior



4.2.1 Parámetros instrumentales del espectrofotómetro. Entre los parámetros instrumentales modificables se encuentran el flujo de Helio, el cual como gas noble y liviano resulta apropiado como portador de la señal. Además, se utiliza para purgar la celda fotoacústica de gases atmosféricos tales como vapor de agua y dióxido de carbono; la ganancia, la cual garantiza una adecuada amplificación de la señal; la velocidad, que se refiere a la rapidez con la que se mueve el espejo del interferómetro de Michelson del espectrofotómetro, determinando así, la frecuencia de modulación del haz incidente en la muestra y finalmente, el número de barridos que indica la cantidad de espectros registrados para posteriormente promediarse.

Estos parámetros fueron ajustados con el propósito de garantizar una alta razón señal a ruido SN . Se fijó un flujo de helio de $10 \text{ cm}^3/\text{s}$ y un factor de amplificación de 10 (nivel de ganancia 3 en el dispositivo usado), se variaron la velocidad y el número de barridos. La razón SN para las diferentes configuraciones se calculó mediante la ecuación (17). [12]

$$SN = \frac{\textit{Media}}{\textit{Desviación estandar}} \quad (17)$$

La desviación estándar, para un nivel de confianza del 99%, se puede estimar dividiendo por cinco, la diferencia entre la señal máxima y la mínima en una región que contenga el fondo espectral. [12]. Así, la expresión (17) llega a ser:

$$SN = \frac{\textit{Media}}{((\textit{Señal máxima} - \textit{Señal mínima})/5)} \quad (18)$$

Para el cálculo de la relación SN se utilizó la región entre 1800 y 2000 cm^{-1} en la cual el espectro $FT-IR$ PAS del carbón no posee señales asociadas a grupos funcionales. [6]

Para la muestra “*Primavera*” se calculó la relación SN , cuyos valores son mostrados a continuación.

Tabla 3. Valores de la relación *SN* de espectros de muestras de carbón de la mina *Primavera* con distintos parámetros del equipo.

Velocidad (cm/s)	<i>SN</i>	Número de Barridos	<i>SN</i>
0,1581	24,1490	32	19,6774
0,3165	25,6146	64	21,4296
0,4747	21,6726	128	22,5740
0,6329	18,0787	256	24,7248
0,9494	18,7389	512	23,4049

De esta forma, atendiendo al mayor valor *SN*, se estableció una velocidad del espejo de 0.3165 (cm/s) y un número de barridos de 256 para registrar los espectros de todas las muestras.

4.2.2. Registro y pretratamiento de espectros. Se registró el espectro *FTIR-PAS* para las 28 muestras de carbón en un rango entre 400 y 4000 cm^{-1} , ya que este posee la mayor información respecto a grupos funcionales del carbón, por ejemplo, grupos alifáticos y aromáticos en regiones cercanos a los 3600 cm^{-1} , 2880 cm^{-1} y 1650 cm^{-1} [13]. Los espectros se auto escalaron, es decir, se calcularon los valores estandarizados (z-score). No se aplicaron procesos de corrección de línea base ni filtrado. Las Figura 5 y Figura 6 muestran los espectros antes y después del proceso de autoescalado [14].

Figura 5. Espectros *FTIR-PAS* tomados de las 28 muestras de carbón.

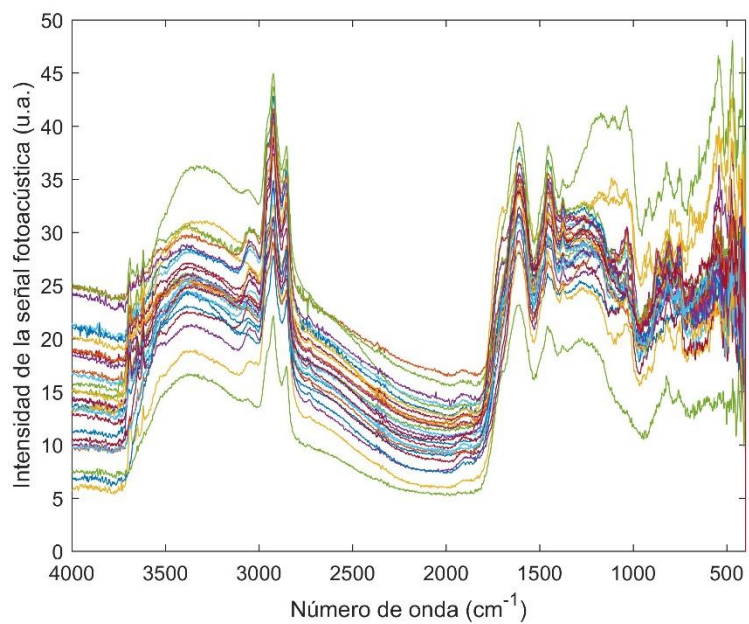
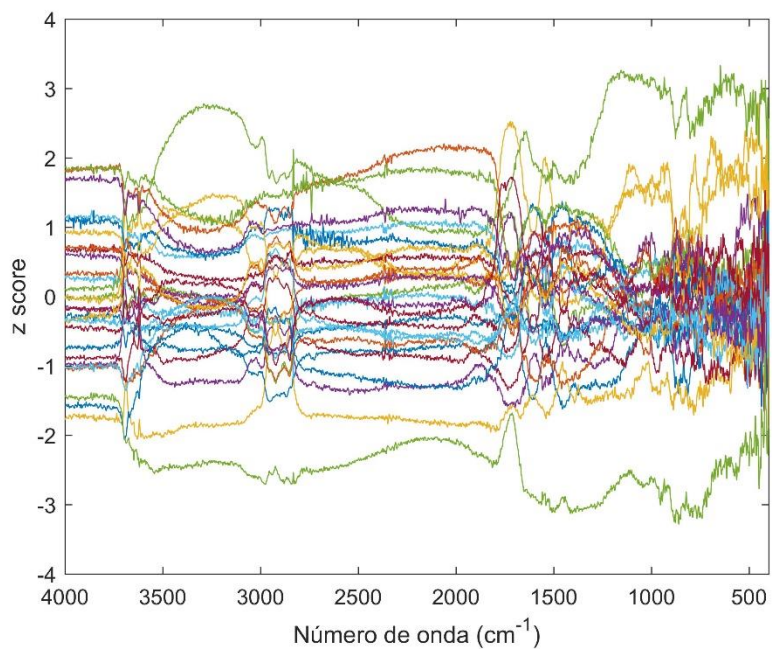


Figura 6. Espectros *FTIR-PAS* normalizados por z-score de las 28 muestras de carbón.



5. ANÁLISIS Y RESULTADOS

5.1. CALCULO DE MODELOS POR *PLS-R*

5.1.1. Determinación de la complejidad y selección del modelo. En primer lugar, se organizó la matriz de espectros $X_{m \times n}$ y el vector de cada propiedad fisicoquímica $\hat{y}_{m \times 1}$, con las cuales se calculó cada modelo. En este caso $m=28$, el número de muestras utilizadas y $n = 7469$, el número de variables independientes (número de onda). Se aplicó el método iterativo *SIMPLS*, explorando modelos con complejidad desde 1 hasta 15 componentes principales.

Para un número de componentes dado se exploraron mil modelos utilizando la estrategia de validación cruzada por *Montecarlo*, manteniendo el 40% del número total de muestras para predicción y las muestras restantes para calibración. [15] [16] [17]

La complejidad de un modelo predictivo hace referencia al número de componentes principales o variables latentes. A diferencia del procedimiento común, que consiste en seleccionar el número de componentes de acuerdo al menor valor promedio de la raíz del error cuadrático medio de predicción (*Root mean square error of prediction, RMSEP*); es posible abordar el problema de la complejidad, utilizando un criterio que se apoya en el análisis del comportamiento del error cuadrático medio de predicción y calibración (*Root mean square error of calibration, RMSECal*) en relación con el error de **sustitución**. La estrategia de sustitución consiste en utilizar todas las muestras para calibrar y todas para validar.

En concreto, cuando se enfrentan para cada modelo obtenido, los errores de calibración y predicción, se encuentran curvas de tendencia como las mostradas en la Figura 7 (a-e). Cada modelo en este espacio coordenado corresponde a un punto y en conjunto se disponen en sectores aproximadamente bisectados por la recta de

identidad, cuyas intersecciones coinciden con la magnitud del error producido por la estrategia de sustitución.

Ahora bien, para determinar el número de componentes consideramos el cociente entre la dispersión de los errores (definida como la longitud del sector), respecto al error de sustitución. Para cada propiedad, se muestra el comportamiento del cociente contra el número de componentes, ver Figura 8 (a-e). El número de componentes seleccionados, inicialmente se establece de acuerdo al mínimo del cociente. Sin embargo, esta cantidad puede ser ajustada conforme se reúnan algunas condiciones que se mencionan a continuación.

En la selección del modelo es importante comprometer *parsimonia* y *optimismo*. La primera se relaciona con modelos de ‘buen comportamiento’, porque manejan pocas variables latentes (pero con el riesgo de quedar sub-entrenado). El optimismo, contrariamente, se asocia con la exagerada especialización (memorización) o sobre ajuste. De manera que resulta conveniente complementar el criterio del párrafo anterior con algunas condiciones que involucren el comportamiento de los errores y el coeficiente de determinación, a saber: $RMSEP > RMSECa$ y $R^2 \geq u$, donde R^2 es el coeficiente de determinación para las muestras de predicción y u un umbral apropiado.

El modelo seleccionado, además de cumplir con todas las condiciones anteriores, será aquel cuya distancia euclidiana a la recta de identidad sea la menor. Recordemos que sobre la recta de identidad el error es cercano al error de la estrategia de validación por sustitución, (estrategia supremamente optimista); por lo tanto, seleccionar un modelo cercano a la recta de identidad emula un comportamiento optimista, pero el modelo no necesariamente lo es.

De acuerdo a esta estrategia, la Tabla 4 muestra el número de componentes y el coeficiente de determinación de los modelos seleccionados. Las componentes se encuentran entre seis y ocho y los valores de R^2 se encuentran en el rango entre 0.94-0.95.

Figura 7. RMSEP de calibración contra RMSECal de predicción de las propiedades a) Ceniza y b) Carbón Fijo c) Humedad y d) Materia Volátil del e) Poder calorífico.

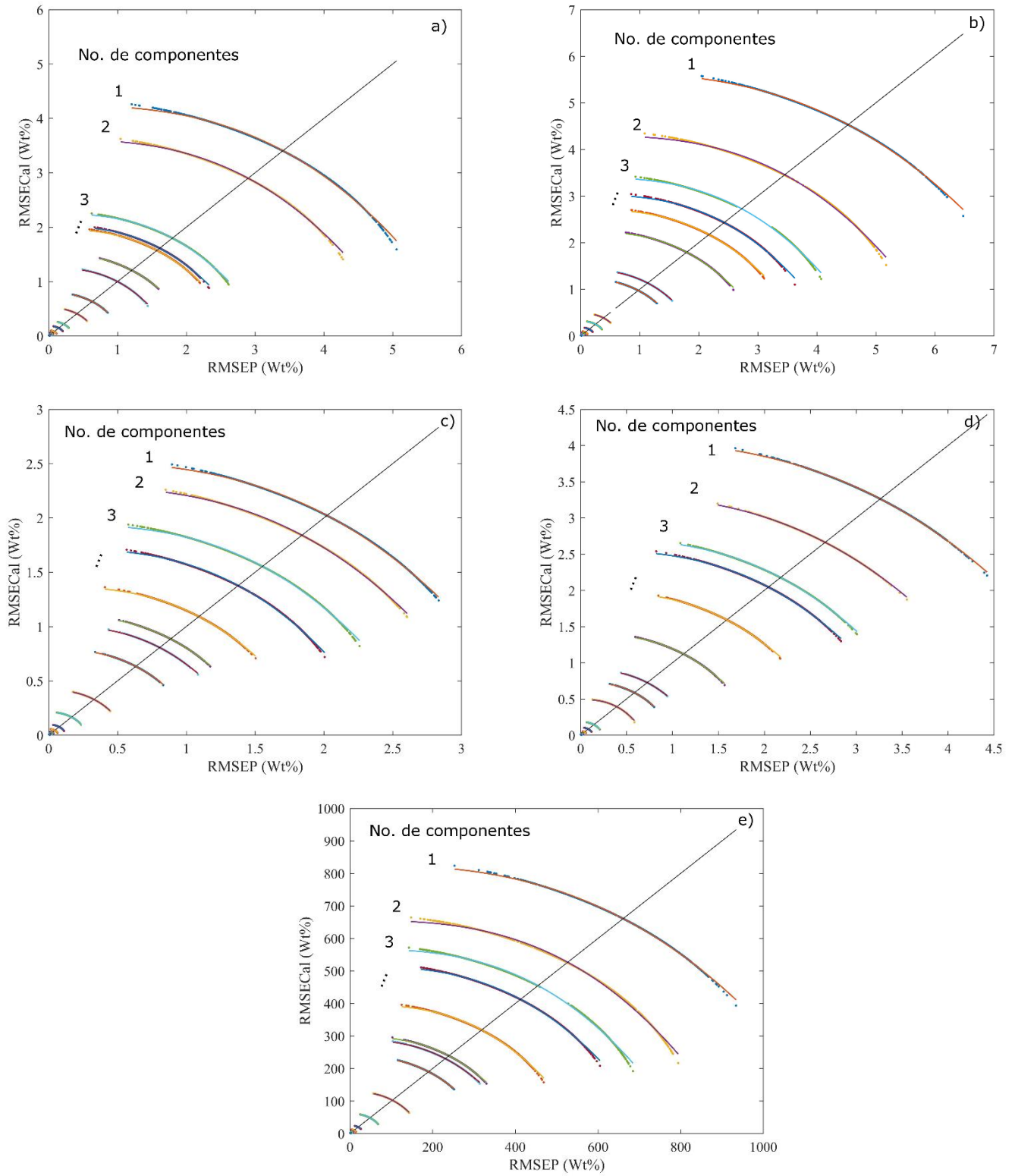


Figura 8. Cociente entre la dispersión de los errores (definida como la longitud del sector), respecto al error de sustitución contra el número de componentes para a) Ceniza y b) Carbón Fijo c) Humedad y d) Materia Volátil del e) Poder calorífico.

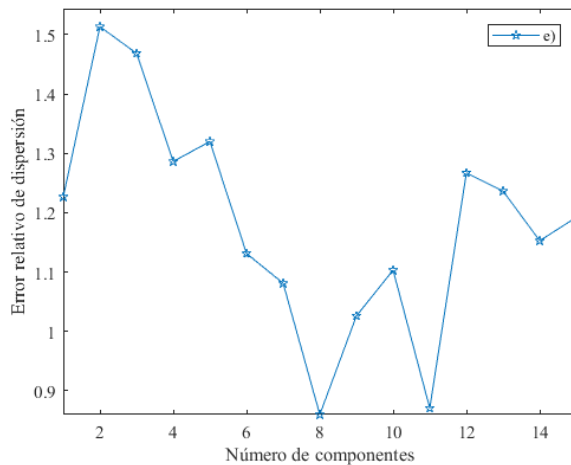
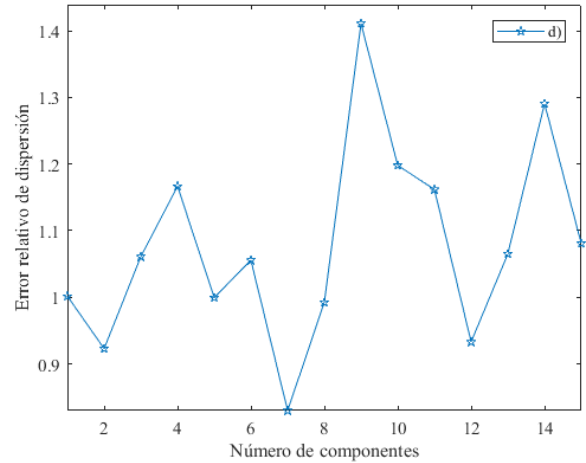
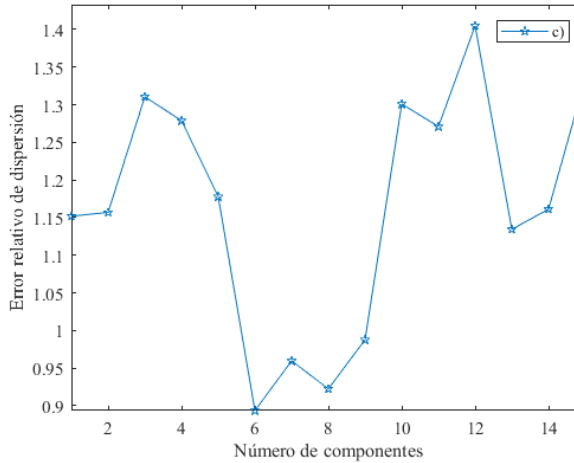
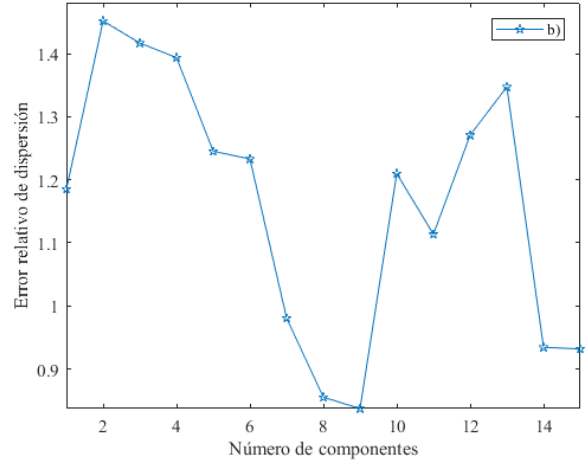
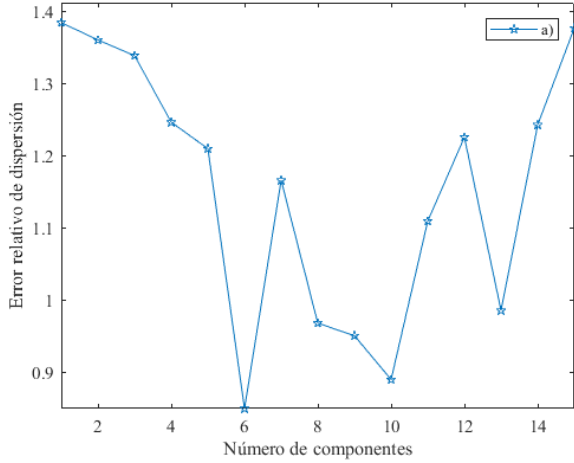


Tabla 4. Número de componentes principales, R^2 , $MSEP$ y $RMSECal$ para cada modelo seleccionado.

Propiedades fisicoquímicas	Número de componentes	R^2 (Predicción)	$RMSEP$ (Wt%)	$RMSECal$ (Wt%)
% Ceniza	6	0.94	1.220	1.190
% Carbón Fijo	8	0.95	0.976	0.975
% Humedad	8	0.94	0.634	0.632
% Materia Volátil	7	0.95	0.720	0.725
Poder calorífico	8	0.95	191.63 (BTU/lb)	189.22 (BTU/lb)

Finalmente, los vectores b o coeficientes del modelo predictivo, se muestran en la Figura 9 (a-e). Nótese que estos vectores oscilan alrededor de cero debido al escalamiento de los espectros (valores z).

5.1.2. Diagnóstico y exactitud de los modelos. Como herramientas de diagnóstico de los modelos construidos se presentan las curvas de valores predichos contra valores medidos, Figura 10, y el comportamiento de los residuales. Por simple inspección de la Figura 10 se verifica la alta correlación entre estos valores, lo cual está de acuerdo con el R^2 estimado y la ausencia de valores atípicos (*outliers*).

Los residuales exhiben un comportamiento *homocedástico*, es decir, no se aprecian tendencias. Así mismo, los histogramas de residuales exhiben una distribución de valores cercanos a cero, ver Figura 11 y Figura 12.

Por otro lado, los errores cuadráticos medio de predicción y calibración tuvieron valores significativamente menores respecto a los valores de las propiedades, excepto para el caso del porcentaje de ceniza y de humedad (ver Tabla 4 y Tabla 5). Esto se debe a que el porcentaje de humedad en el carbón puede variar significativamente por las condiciones ambientales a las que está sometida la

muestra. Así mismo, el porcentaje de ceniza, asociado principalmente al contenido de minerales, presenta una baja correlación con el espectro infrarrojo.

Tabla 5. Valores mínimos, máximo y promedio de las propiedades fisicoquímicas medidas.

Propiedad	Valor mínimo	Valor máximo	Valor promedio
Poder calorífico (BTU/lb)	11495,200	14521,550	13246,681
Carbón Fijo (Wt%)	44,478	63,196	55,264
Materia Volátil (Wt%)	28,640	44,294	36,215
Ceniza (Wt%)	0,6600	16,8458	5,8516
Humedad (Wt%)	0,2696	7,1127	2,6695

Figura 9. Modelos seleccionados para a) Ceniza y b) Carbón Fijo c) Humedad y d) Materia Volátil del e) Poder calorífico.

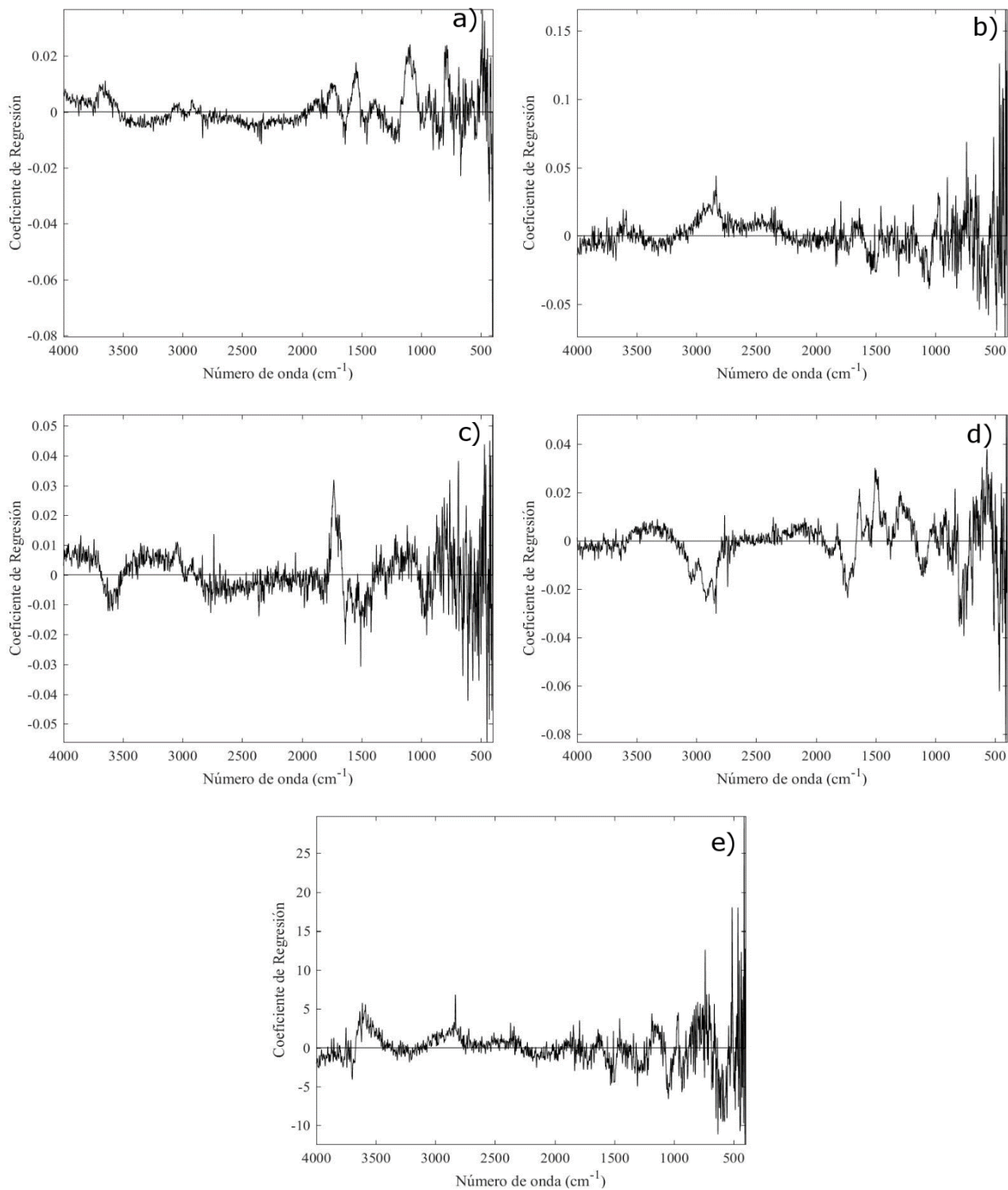
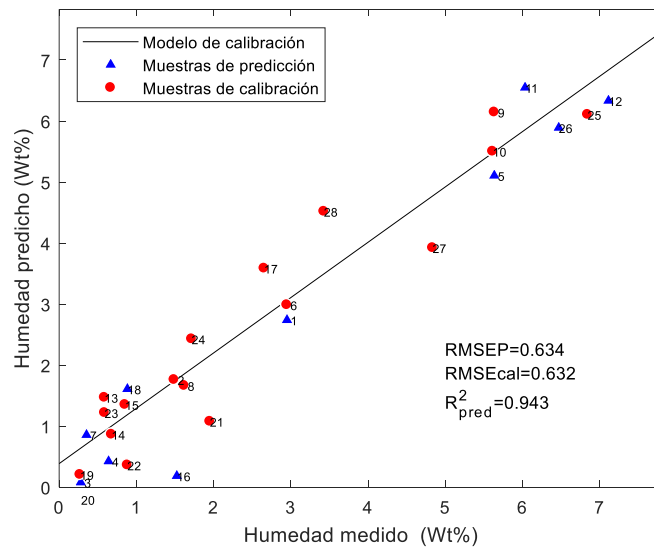
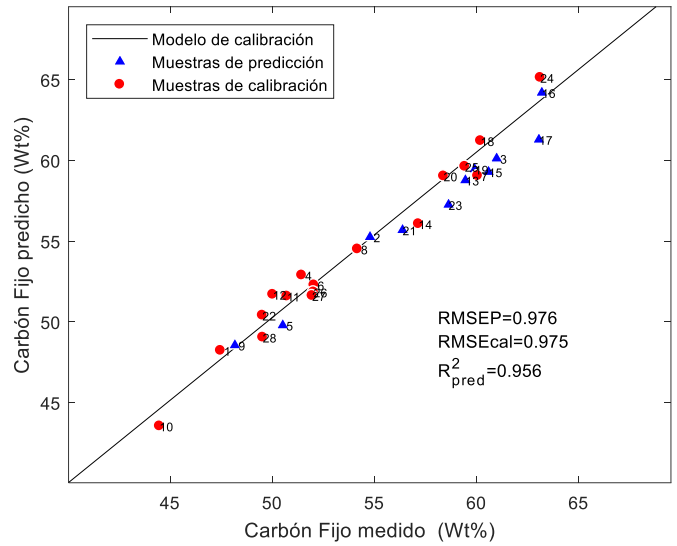
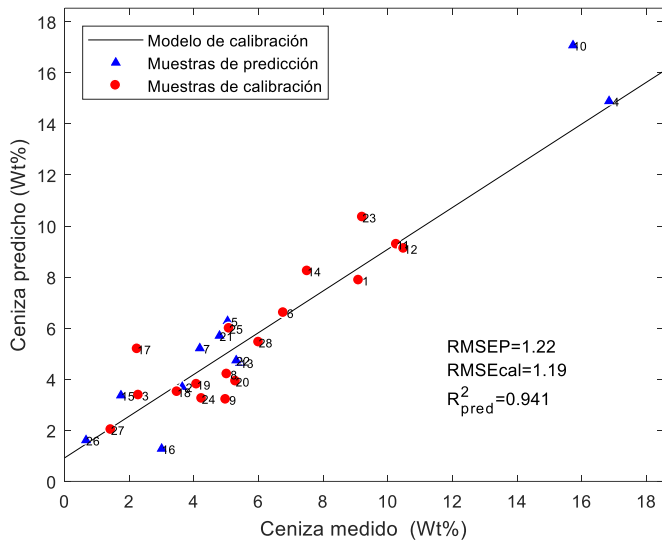


Figura 10. Gráfica de predicciones por los modelos PLS contra los valores medidos para cada propiedad fisicoquímica.



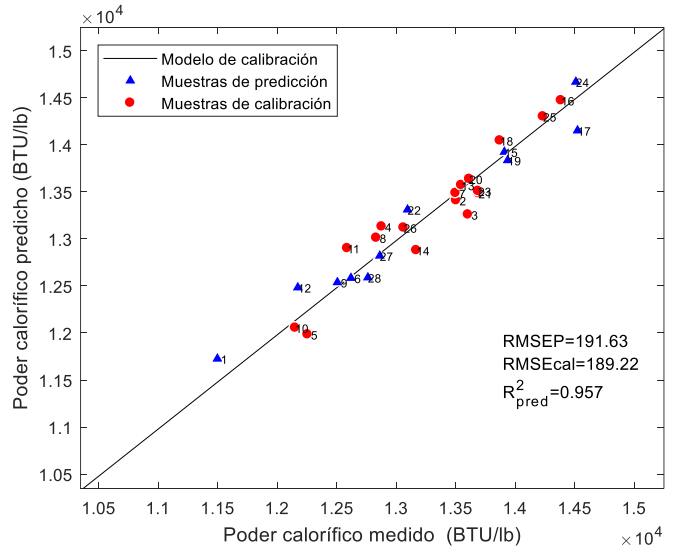
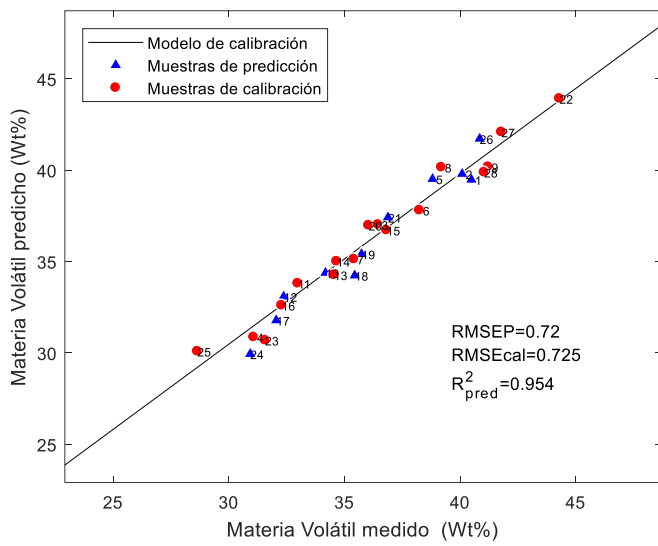
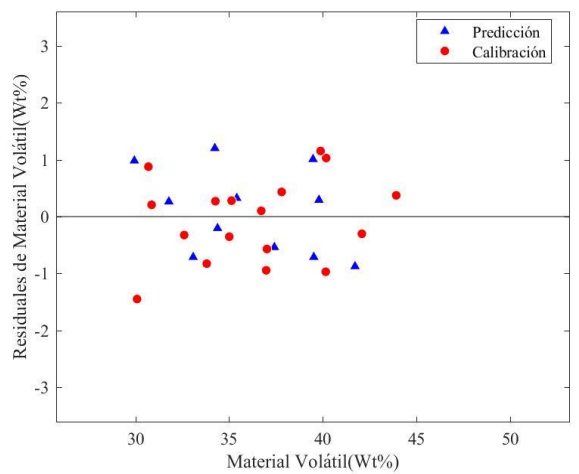
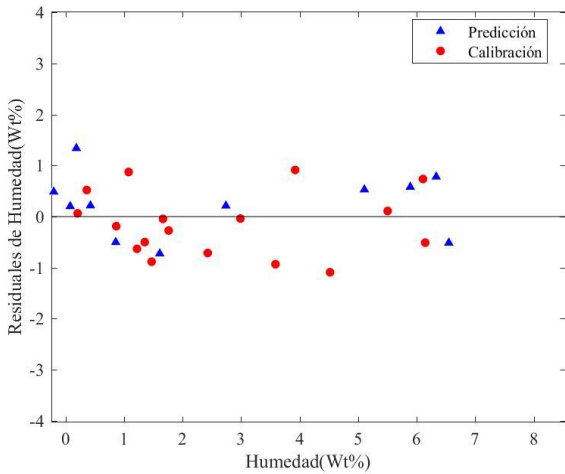
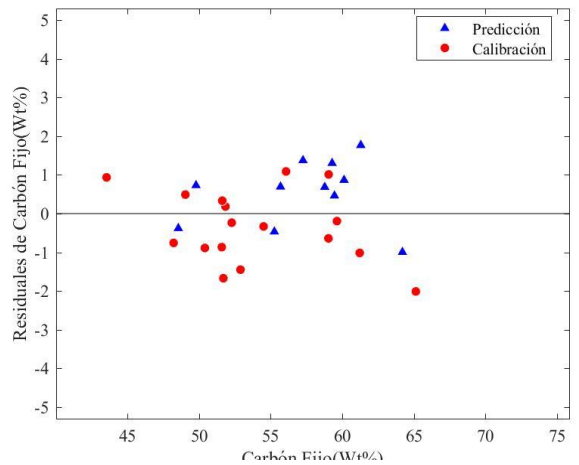
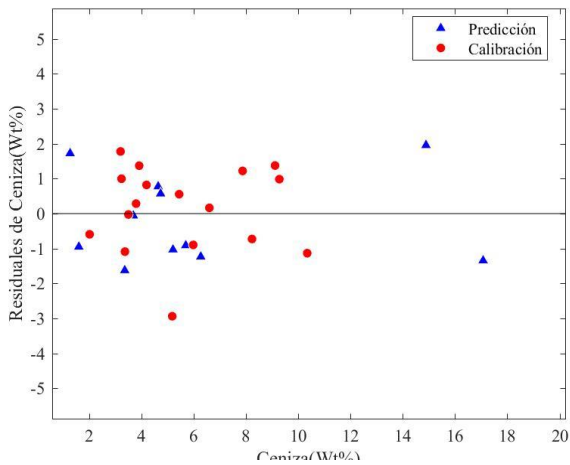


Figura 11. Residuales contra los valores de sus propiedades



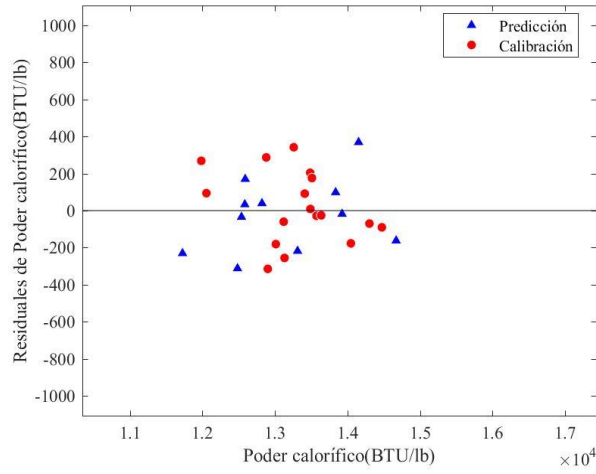
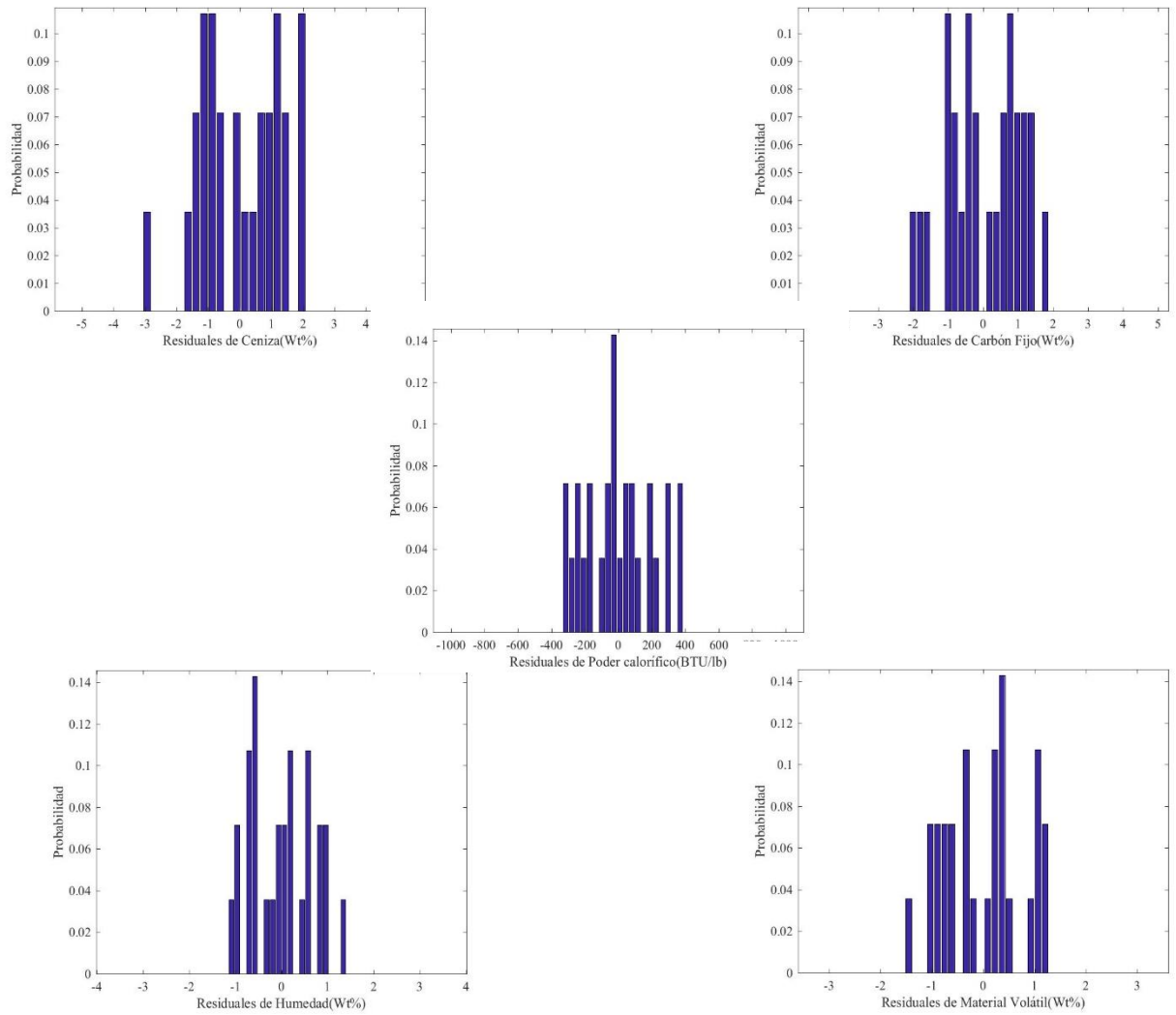


Figura 12. Histograma de los residuos para todas las propiedades.



6. CONCLUSIONES

En este trabajo se construyeron a partir de 28 muestras de carbón, provenientes de diferentes minas del territorio colombiano, modelos de calibración multivariable para la predicción de porcentaje de ceniza, humedad, materia volátil, carbón fijo y poder calorífico, utilizando espectroscopía fotoacústica infrarroja por transformada de Fourier (*FTIR-PAS, Fourier Transform Infrared Photoacoustic spectroscopy*) y regresión de mínimos cuadrados parciales (*PLSR, Partial least square regression*). Se realizó una exploración previa de los parámetros instrumentales: velocidad de barrido, flujo de helio y número de escaneos para el espectrofotómetro *Thermo Scientific Nicolet iS50 FT-IR* acoplado a una cámara fotoacústica (*MTEC* modelo 300).

El diagnóstico de los modelos, apoyado en la raíz del error cuadrático medio y el coeficiente de determinación para las muestras de predicción, da cuenta de un buen resultado, en términos de linealidad, para la predicción de materia volátil, carbón fijo y poder calorífico, y un desempeño moderado para el porcentaje de humedad y ceniza. Estos modelos están de acuerdo con consideraciones *a priori* sobre las correlaciones de las propiedades y los espectros de las muestras. Así mismo, el *comportamiento homocedástico* de los residuales y su *normalidad* (respecto a cero) confirman la ausencia de tendencias conducidas por los datos.

Por otro lado, se propuso una estrategia de validación cruzada basada en el comportamiento de los errores de calibración y predicción (validación) generados por los métodos de *Montecarlo* y *sustitución*. Esto permitió establecer un criterio para la determinación de la complejidad y selección de los modelos.

REFERENCIAS BIBLIOGRÁFICAS

1. **Portafolio, Noticia del.** *Colombia espera producir 85,5 millones de toneladas de carbón este año.* Tomado 9 de enero de 2017. <http://www.portafolio.co/economia/colombia-producira-85-millones-toneladas-carbon-2016-494519>.
2. *Producción de carbón se acercaría a meta de 100 millones de toneladas.* Tomado 9 de enero de 2017. <http://www.portafolio.co/economia/carbon-aumento-su-produccion-501703>.
3. **ANM, Agencia Nacional Minera.** *Carbón.* Tomado el 12 de noviembre de 2017 <https://www.anm.gov.co/sites/default/files/DocumentosAnm/carbon.pdf>
4. **Dr. Cortés, J. Vicente.** *CARBÓN.* Tomado el 12 de noviembre de 2017 <http://www.factoria3.com/documentos/CARBON.pdf>
5. **Randall D. Tobias, SAS Institute Inc., Cary, NC.** *An Introduction to Partial Least Squares Regression.*
6. **Torrado, Camilo Andres Sanchez.** *PREDICCIÓN DE PROPIEDADES FISICOQUÍMICAS DE CARBONES COLOMBIANOS DETERMINADOS POR ESPECTROSCOPIA INFRARROJA (FTIR-ATR) Y MÉTODOS QUIMIOMÉTRICOS.*
7. **García Gómez, Roberto Carlos.** *Espectroscopia Fotoacústica: Una teoría antigua con nuevas aplicaciones.*
8. **D3173, ASTM.** *Standard Test Method for Moisture in the Analysis Sample of Coal and Coke.* Martes Julio 26 08:02:45 2016.
9. **ASTM D3174, Martes Julio 26 07:40:57 2016.** *Standard Test Method for Ash in the Analysis Sample of Coal and Coke from Coal.* s.l. : www.astm.org.
10. **ASTM D3175, Martes Julio 26 08:02:45 2016.** *Standard Test Method for Volatile Matter in the Analysis Sample of Coal and Coke.* s.l. : www.astm.org.
11. **ASTM D5865, Martes Julio 26 07:40:57 2016,.** *Standard Test Method for Gross Calorific Value of Coal and Coke.* s.l. : www.astm.org.
12. **Holler, Douglas A. Skoog y F James.** *Principios de análisis instrumental, sexta edición.* s.l. : Cengage learning.

13. **Ruiz, Jorge Armando Orrego.** *Estudio de la estructura de cinco carbonos Colombianos por espectroscopía fotoacústica en el infrarrojo medio.*
14. **MathWorks-plsregress.** <https://es.mathworks.com/help/stats/plsregress.html>. Tomado el 16 de octubre de 2017.
15. **Peter Filzmoser, Bettina Liebmann ay Jurt Varmuza.** *Repeat double cross Validation.* s.l. : CHEMOMETRICS, 24 diciembre 2008.
16. **Qing-Song Xu a, Yi-Zeng Liang.** *Monte Carlo cross validation.* s.l. : Chemometrics and Intelligent Laboratory Systems, Julio 2000.
17. **Yi Ping Du, Sumaporn Kasemsumran, Katsuhiko Maruo, Takehiro Nakagawa, Yukihiro Ozaki.** *Ascertainment of the number of samples in the validation set in Monte Carlo cross validation and the selection of model dimension with Monte Carlo cross validation.* s.l. : Chemometrics and Intelligent Laboratory, Septiembre 2005.
18. **Otto, Matthias.** *Chemometrics, Statistics anda computer Application in Analitical Chemistry.*
19. **UPME, Unidad de planeación minero energética.** *LA CADENA DEL CARBÓN.*
20. **Marek W. Urban, Scott R. Gaboury, William F. McDonald, and Ann M. Tiefenthaler.** *Probing Polymer Structures by Photoacoustic Fourier Transform Infrared Spectroscopy.*
21. **Deepak V. Bageshwar*, Avinash S. Pawar, Vineeta V. Khanvilkar, Vilasrao J. Kadam, Eurasian J. Anal. Chem.** *Photoacoustic Spectroscopy and Its Applications– A Tutorial Review.*
22. **Zongyu Hou, Zhe Wang¹, Tingbi Yuan¹, Jianmin Liu, Zheng Li, Weidou Ni.** *A Hybrid Quantification Model and Its Application for Coal Analysis Using Laser Induced Breakdown Spectroscopy.* Journal of Analytical Atomic Spectrometry.
23. **J.F. McClelland, S.J.Bajic, R.W.Jones and L.M.Seaverson.** *Introduction to Photoacoustic Spectroscopy with Step Scan and Constant Velocity Scan FTIR Spectrometers.* s.l. : Ames Laboratory.
24. **András Miklós, Stefan Schäfer and Peter.** *PHOTOACOUSTIC SPECTROSCOPY, THEORY.*

25. **LOWRY, D. WARREN VIDRINE and S. R.** *Photoacoustic Fourier Transform IR Spectroscopy and Its Application to Polymer Analysis.*
26. **Abdi, Hervé.** *Partial Least Square Regression PLS-Regression.* s.l. : The University of Texas at Dallas,.
27. **Jong, Sijmen tie.** *SIMPLS: an alternative approach to partial least squares regression.* s.l. : Elsevier Science Publishers B.V., Amsterdam.
28. **Miller, James N. Miller Jane C.** *Estadística y quimiometría para química analítica.*
29. **ASTM D7582, Martes Julio 26 07:51:075 2016.** *Standard Test Methods for Proximate Analysis of Coal and Coke by Macro Thermogravimetric Analysis.* s.l. : www.astm.org.
30. **Ferre´, Nicolaas (Klaas) M. Faber and Joan.** *On the numerical stability of two widely used PLS algorithms.* s.l. : Chemometrics, 29 November 2007.
31. **Rodrigo R. de Oliveira, Luiz S. das Neves and Kássio M. G. de Lima.** *Experimental Desing, Near-Infrared Spectroscopy, and Multivariate Calibration: An Advanced Project in a Chemometrics Course.* s.l. : CHEMICALEDUACATION.
32. **Lesly V. Melendez, Adriana Lanche, Jorge A. Orrego Ruiz, Enrique Mejia Ospino.** *Prediction of the SARA analysis of colombian crude oils using ATR-FTIR spectroscopy and chemometrics methods.* s.l. : Journal of petroleum science and engineering.

BIBLIOGRAFÍA

Abdi, Hervé. Partial Least Square Regression PLS-Regression. s.l. : The University of Texas at Dallas,.

ANM, Agencia Nacional Minera. Carbón. Tomado el 12 de noviembre de 2017 <https://www.anm.gov.co/sites/default/files/DocumentosAnm/carbon.pdf>

ASTM D3174. Martes Julio 26 07:40:57 2016. Standard Test Method for Ash in the Analysis Sample of Coal and Coke from Coal. s.l. : www.astm.org.

ASTM D3175. Martes Julio 26 08:02:45 2016. Standard Test Method for Volatile Matter in the Analysis Sample of Coal and Coke. s.l. : www.astm.org.

ASTM D5865. Martes Julio 26 07:40:57 2016,. Standard Test Method for Gross Calorific Value of Coal and Coke. s.l. : www.astm.org.

ASTM D7582. Martes Julio 26 07:51:075 2016. Standard Test Methods for Proximate Analysis of Coal and Coke by Macro Thermogravimetric Analysis. s.l. : www.astm.org.

D3173, ASTM. Standard Test Method for Moisture in the Analysis Sample of Coal and Coke. Martes Julio 26 08:02:45 2016.

García Gómez, Roberto Carlos. Espectroscopia Fotoacústica: Una teoría antigua con nuevas aplicaciones.

LOWRY, D. WARREN VIDRINE and S. R. Photoacoustic Fourier Transform IR Spectroscopy and Its Application to Polymer Analysis.

Peter Filzmoser, Bettina Liebmann ay Jurt Varmuza. Repeat double cross Validation. s.l. : CHEMOMETRICS, 24 diciembre 2008.

Qing-Song Xu a, Yi-Zeng Liang. Monte Carlo cross validation. s.l. : Chemometrics and Intelligent Laboratory Systems, Julio 2000.

Randall D. Tobias, SAS Institute Inc., Cary, NC. An Introduction to Partial Least Squares Regression.

Ruiz, Jorge Armando Orrego. Estudio de la estructura de cinco carbones Colombianos por espectroscopía fotoacústica en el infrarrojo medio.

Torrado, Camilo Andres Sanchez. PREDICCIÓN DE PROPIEDADES FISICOQUÍMICAS DE CARBONES COLOMBIANOS DETERMINADOS POR ESPECTROSCOPIA INFRARROJA (FTIR-ATR) Y MÉTODOS QUIMIOMÉTRICOS.

Yi Ping Du, Sumaporn Kasemsumran, Katsuhiko Maruo, Takehiro Nakagawa, Yukihiro Ozaki. Ascertainment of the number of samples in the validation set in Monte Carlo cross validation and the selection of model dimension with Monte Carlo cross validation. s.l. : Chemometrics and Intelligent Laboratory, Septiembre 2005.