

Partial least squares (PLS) regression and its application to coal analysis

Carlos E. Alciaturi¹, Marcos E. Escobar^{1,2}, Carlos De La Cruz³ and Carlos Rincón⁴

¹Instituto Zuliano de Investigaciones Tecnológicas (INZIT-CICASI), Apartado Postal 1114. Maracaibo, Venezuela. Tel. 58-261-7910952 Fax 58-261-7913769. E-mail: alciaturi@hotmail.com.

²Postgrado en Geología Petrolera, División de Estudios para Graduados, Facultad de Ingeniería.

³Laboratorio de Espectroscopía Molecular y Atómica, Departamento de Química, Facultad Experimental de Ciencias.

⁴Escuela de Ingeniería Química, Facultad de Ingeniería, Universidad del Zulia. Maracaibo, Venezuela

Abstract

Instrumental chemical analysis methods use the relationships between a signal obtained and a property (generally a concentration) of the system under study. The study and applications of these relations is known as chemometrics, a discipline of intense development, with ample applications in chemical and process industry and in environmental studies. The method of partial least squares (PLS) is one of the most used in chemometrics. This method is closely related to principal components regression (PCR). PLS have theoretical and computational advantages that have led to a great number of applications. The numbers of Internet sites referring to PLS are hundreds of thousands. Here, we give the fundamentals and show an application to prediction of coal properties from mid-infrared data, with the purpose of developing fast, non-destructive methods of analysis for these materials.

Key words: PLS, multivariate regression, chemometrics, coal analysis.

La regresión de mínimos cuadrados parciales (PLS) y su aplicación al análisis del carbón mineral

Resumen

Los métodos instrumentales de análisis químico hacen uso de las relaciones entre la señal obtenida y una propiedad del sistema estudiado (generalmente, una concentración). Los avances en electrónica y computación han hecho posible un rápido progreso en la adquisición de datos y en su transmisión y procesamiento. La aplicación de diversos métodos matemáticos al cálculo de concentraciones y otras propiedades a partir de datos instrumentales se conoce como **quimiometría** y es un área de intensa actividad, por sus amplias aplicaciones en la industria química, de procesos y en estudios ambientales. Uno de los métodos más usados en quimiometría es el método de mínimos cuadrados parciales, conocido por sus iniciales en inglés, PLS ("partial least squares"). Este método, relacionado con la regresión de componentes principales, PCR ("principal components regression") posee ventajas teóricas y computacionales que han llevado a innumerables aplicaciones. Se encuentran en Internet decenas de miles de referencias solamente para el PLS lineal. En este artículo, se explica los fundamentos del método y se muestra una aplicación a la predicción de propiedades de carbones minerales a partir de datos del infrarrojo medio, con el objetivo de desarrollar métodos de análisis rápidos y no destructivos para estos materiales.

Palabras clave: PLS, regresión multivariada, quimiometría, análisis del carbón mineral.

Introducción

Es frecuente, para medir propiedades de sistemas físicos, el hacer uso de señales que puedan ser logradas de manera rápida y eficiente, por instrumentos electrónicos o electro-ópticos. La propiedad que interesa debe estar relacionada de alguna manera con la intensidad de la señal obtenida. El estudio de estas relaciones es conocido como quimiometría, y es un área de rápido desarrollo, gracias a la disponibilidad de computadoras, espectrofotómetros, sensores electroquímicos, y otros que permiten realizar análisis rápidos, económicos y no destructivos. El trabajo pionero en PLS fue realizado por H. Wold en los años sesenta en el área de econometría [1]. S. Wold, H. Martens y B. Kowalski estuvieron entre los primeros en aplicar este método en química [1-4]. El método se propone identificar los factores (combinaciones lineales de las variables independientes) que mejor modelan las variables dependientes. Es posible manejar eficientemente conjuntos de datos donde existen muchas variables fuertemente correlacionadas y cantidades importantes de ruido aleatorio. Se ha demostrado que este método es una buena alternativa a los métodos más antiguos de regresión lineal múltiple (MLR) porque es "robusto". Esto significa que el modelo matemático no es muy alterado cuando se toman en cuenta nuevas muestras en la calibración. Entre sus ventajas se encuentran: análisis rápidos y no destructivos de muestras; determinación simultánea de varias propiedades; automatización de las medidas; y detección de "outliers" (muestras diferentes a las utilizadas en la calibración). Sin embargo, existe la debilidad, común a todos los métodos quimiométricos, de ser correlativos y no causales. Esto significa que las relaciones determinadas serán aplicables únicamente a muestras similares (cuyas propiedades estén dentro del mismo rango) a las usadas para la creación del modelo.

El análisis multivariado

En la literatura se encuentran numerosas referencias a métodos de análisis multivariado (por ejemplo, MLR (regresión lineal múltiple), PCR (regresión de componentes principales) y PLS (mínimos cuadrados parciales). Geladi y Kowalski publicaron un artículo de revisión so-

bre el método PLS [1-2]. Además, recientemente Wold y Martens publicaron artículos donde se relata el desarrollo temprano del método [3, 4]; y Wold *et al.* analizaron tendencias recientes [5]. Artículos de revisión bibliográfica sobre quimiometría se publican regularmente en la revista *Analytical Chemistry* (ver, por ej., [6]). En la misma revista se han publicado reportes sobre las aplicaciones del análisis multivariado [7-9]. Asimismo, existen textos sobre el tema [10-12]. En estos métodos, se consideran dos grupos de variables, llamadas variables independientes (representadas por una matriz \mathbf{X}) y variables dependientes (representadas por la matriz \mathbf{Y}). Tomando un conjunto de muestras, a las cuales se les determina los valores de \mathbf{X} y \mathbf{Y} (por ejemplo, \mathbf{X} puede representar las intensidades de absorción medidas por un espectrofotómetro, y \mathbf{Y} las concentraciones de ciertos componentes, determinadas por gravimetría, titulación u otro método) se trata de establecer un modelo (generalmente lineal) $\mathbf{Y} = f(\mathbf{X})$. Este conjunto de muestras es llamado conjunto de calibración o de entrenamiento.

Se considera el caso de un modelo lineal con una variable dependiente, representado por la Ecuación 1 (donde las letras mayúsculas representan matrices, y las minúsculas vectores):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{f} \quad (1)$$

y representa la variable dependiente, \mathbf{X} representa las variables independientes, β es el vector solución y \mathbf{f} el vector *error* o *residual*. Esta ecuación se muestra gráficamente en la Figura 1. Los parámetros del modelo (contenidos en β) son llamados coeficientes de regresión o sensibilidades.

Es frecuente, cuando los datos son obtenidos por métodos instrumentales en química analítica, encontrar un mayor número de variables independientes que muestras; lo cual puede resultar en infinitas soluciones para el sistema en la Ecuación 1. Por ejemplo, los avances de la electrónica permiten medir y registrar la absorbancia a un número muy grande de longitudes de onda. Esto se da en espectroscopia de NIR (infrarrojo cercano), FT-IR (infrarrojo por transformada de Fourier) o NMR (resonancia magnética nuclear). Tales espectros digitalizados pueden dar 1000-2000 variables por muestra.

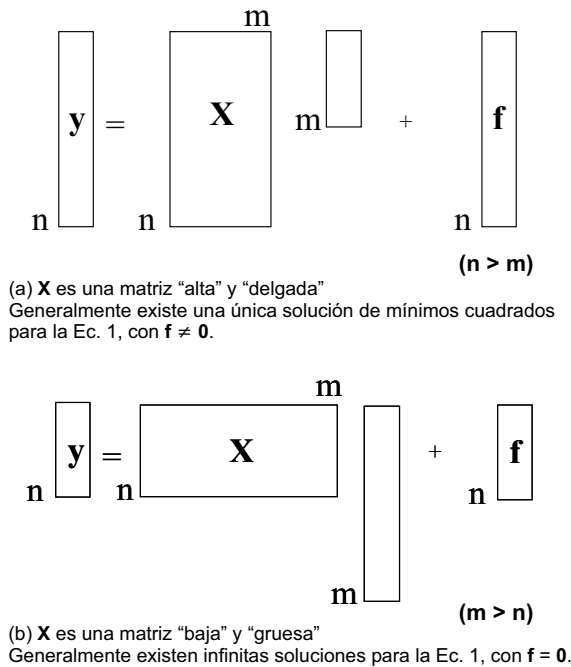


Figura 1. Representación gráfica de la ecuación 1.

Este problema puede ser enfrentado de varias maneras: selección de variables (escoger las variables más relevantes); reducción de dimensionalidad de la matriz \mathbf{X} calculando los componentes principales o variables latentes (PCR y PLS [1-12]) y/o transformación de las variables (por ej. mediante la transformada de Fourier). Todos estos métodos resultan en una matriz \mathbf{X} de menor dimensión, lo cual facilita los cálculos y hace que el sistema sea más "robusto".

Se acostumbra *validar* el modelo obtenido utilizando un nuevo conjunto de muestras (llamado conjunto de predicción o de prueba) para el cual se miden los valores de las variables independientes y estas variables son utilizadas, junto con los coeficientes de regresión, para predecir los valores de las variables dependientes. Un método de validación de uso frecuente, es la *validación cruzada*, donde una fracción de las muestras de referencia es omitida al construir el modelo [1, 2]. La fracción omitida puede variar entre una ("bootstrapping") y la mitad del total de muestras. El método de validación cruzada da una estimación más realista de las desviaciones que pueden esperarse al aplicar el modelo a nuevas muestras.

El Método de Mínimos Cuadrados Parciales (PLS)

La regresión de mínimos cuadrados parciales (PLS) es el "caballo de batalla" de la quimiometría. Es un método relacionado con el análisis de componentes principales (PCA) [1]. Estos métodos tienen ventajas intrínsecas cuando se le compara con métodos univariados. Todas las variables relevantes son incluidas en el modelo PLS. En lo que sigue, se hará referencia al llamado modelo PLS1, donde se considera una variable dependiente, representada por el vector \mathbf{y} . Cuando se consideran varias variables dependientes, el modelo es denominado PLS2, y las variables se representan por una matriz \mathbf{Y} . La suposición básica de todos estos modelos es que el sistema o proceso estudiado depende de un número pequeño de *variables latentes* (V.L.). Este concepto es similar al de componentes principales. Las variables latentes son estimadas como combinaciones lineales de las variables observadas, como se explica más adelante.

En los modelos PLS, se establece una representación de la matriz \mathbf{X} en término de estas V.L.:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{2}$$

donde \mathbf{T} representa "scores" (término que puede ser traducido como "resultados"); y \mathbf{P}^T "loadings" (término que puede ser traducido como "cargas"). La matriz \mathbf{X} queda descompuesta en un número de "variables latentes", cada una caracterizada por un vector \mathbf{t} y un vector \mathbf{p}^T .

De esta forma, es posible representar una matriz \mathbf{X} (dentro de un cierto error) por una matriz \mathbf{T} con un número menor de columnas. Esta descomposición se muestra en la Figura 2. Si se incluyen todas las variables latentes, el error es cero ($\mathbf{E} = \mathbf{0}$). El modelo PLS1 se desarrolla de modo que las primeras variables latentes ($\mathbf{t}_1, \mathbf{t}_2, \dots$) son las más importantes para explicar el vector " \mathbf{y} " en el conjunto de calibración. El número de variables latentes necesarias para explicar la matriz \mathbf{X} (dentro de un error máximo permitido) es una medida de la complejidad del modelo. Otros vectores calculados durante la etapa de calibración son \mathbf{w}^T ("pesos" en el espacio de \mathbf{X}), y \mathbf{b} ("sensibilidades"). El algoritmo para la calibra-

$$\begin{aligned}
 \begin{matrix} m \\ \boxed{\mathbf{X}} \\ n \end{matrix} &= \begin{matrix} a \\ \boxed{\mathbf{T}} \\ n \end{matrix} \begin{matrix} m \\ \boxed{\mathbf{P}^T} \\ a \end{matrix} + \begin{matrix} m \\ \boxed{\mathbf{E}} \\ n \end{matrix} \\
 &= \begin{matrix} \boxed{\mathbf{t}_1} \\ n \end{matrix} \begin{matrix} \boxed{\mathbf{p}_1^T} \\ m \end{matrix} + \begin{matrix} \boxed{\mathbf{t}_2} \\ n \end{matrix} \begin{matrix} \boxed{\mathbf{p}_2^T} \\ m \end{matrix} \\
 &+ \dots + \begin{matrix} \boxed{\mathbf{t}_a} \\ n \end{matrix} \begin{matrix} \boxed{\mathbf{p}_a^T} \\ m \end{matrix} + \begin{matrix} m \\ \boxed{\mathbf{E}} \\ n \end{matrix}
 \end{aligned}$$

Figura 2. La matriz **X** es descompuesta en "scores" (**t**) y "loadings" (**p**). La matriz **T** contiene un número **a** de scores, con **a** < **m**.

ción (nonlinear iterative partial least squares, NI-PALS), [1-2] se muestra en la Tabla 1.

La relación entre **y** y **T** es:

$$\mathbf{y} = \mathbf{T}\mathbf{b} + \mathbf{f} \tag{3}$$

donde **b** se calcula de modo de minimizar **f**, el vector error en esta expresión.

Los vectores **w^T**, **p^T**, y **b** son necesarios para predecir la propiedad dependiente **y** para un nuevo conjunto de muestras. En ese caso, se determina experimentalmente **X** y se calcula (utilizando los vectores **w'** y **p'**) la matriz **T**. El vector **y** es estimado usando los coeficientes **b** previamente calculados durante la calibración:

$$\mathbf{y}_{pred} = \mathbf{T}\mathbf{b} \tag{4}$$

El algoritmo para la predicción se muestra en la Tabla 2; ver también [1, 2, 10-12].

Como se observa en la Tabla 1, este algoritmo produce una deflación de la matriz **X** y el vector **y**. Los vectores y matrices calculados por este algoritmo tienen las siguientes propiedades:

- Los **t** ("scores") son vectores ortogonales entre sí [1].
- Los **w^T** ("weights") son ortogonales entre sí [1].
- El residual **f** es ortogonal a los **t** previos [13].
- El número máximo de variables latentes no excede el rango de la matriz **X** [13].

Tabla 1
El algoritmo PLS1 (calibración)

X y el vector **y** son previamente centrados y escalados a varianza 1

$$\mathbf{f}_0 = \mathbf{y}; \mathbf{E}_0 = \mathbf{X}$$

Para la iteración **h** (con **h** = 1, 2, ..., **a**, donde **a** es el número de variables latentes):

- (1) $\mathbf{w}_h^T = \mathbf{f}_{h-1}^T \mathbf{E}_{h-1} / \mathbf{f}_{h-1}^T \mathbf{f}_{h-1}$
- (2) $\mathbf{w}_h^T = \mathbf{w}_h^T / \|\mathbf{w}_h^T\|$ (normalización)
- (3) $\mathbf{t}_h = \mathbf{E}_{h-1} \mathbf{w}_h / \mathbf{w}_h^T \mathbf{w}_h$
- (4) $\mathbf{p}_h^T = \mathbf{t}_h^T \mathbf{E}_{h-1} / \mathbf{t}_h^T \mathbf{t}_h$
- (5) $\mathbf{t}_h = \mathbf{t}_h / \|\mathbf{p}_h^T\|$
- (6) $\mathbf{w}_h^T = \mathbf{w}_h^T / \|\mathbf{p}_h^T\|$
- (7) $\mathbf{p}_h^T = \mathbf{p}_h^T / \|\mathbf{p}_h^T\|$ (normalización)
- (8) $\mathbf{b}_h = \mathbf{f}_h^T \mathbf{t}_h / \mathbf{t}_h^T \mathbf{t}_h$
- (9) $\mathbf{E}_h = \mathbf{E}_{h-1} - \mathbf{t}_h \mathbf{p}_h^T$
(Cálculo de los residuales de **X**)
- (10) $\mathbf{f}_h = \mathbf{f}_{h-1} - \mathbf{b}_h \mathbf{t}_h$
(Cálculo de los residuales de **y**)
- (11) **h** = **h** + 1

De aquí se vuelve al paso (1)

Tabla 2
El algoritmo PLS1 (predicción)

Dado un conjunto de muestras, la matriz **X** es previamente centrada y escalada, utilizando los parámetros de promedio y desviación estándar calculados en la etapa de calibración.

Asimismo, en los siguientes pasos, se usan los valores de **w'**, **p'**, y **b** encontrados en la calibración

$$\mathbf{E}_0 = \mathbf{X}$$

Para la iteración **h** (con **h** = 1, 2, ..., **a**):

- (1) $\mathbf{t}_h = \mathbf{E}_{h-1} \mathbf{w}_h / \mathbf{w}_h^T \mathbf{w}_h$
- (2) $\mathbf{E}_h = \mathbf{E}_{h-1} - \mathbf{t}_h \mathbf{p}_h^T$
- (3) **h** = **h** + 1

De aquí se vuelve al paso (1)

El vector **y** es estimado como sigue:

$$\mathbf{y} = \sum \mathbf{b}_h \mathbf{t}_h = \mathbf{T}\mathbf{b}$$

donde **b_h** es un escalar, y **h** = 1, 2, ..., **a**.

- La aproximación obtenida cuando se consideran todas las variables latentes es idéntica a la solución por la matriz pseudo-inversa [14].

Debe agregarse que es posible calcular las matrices indicadas anteriormente sin utilizar el algoritmo NIPALS [15-17]. También es posible representar el modelo PLS calculando un vector de coeficientes β (ver Ec. 1) como se indica a continuación [18]. Defínase una matriz \mathbf{R} como:

$$\mathbf{R} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \quad (5)$$

La matriz de scores \mathbf{T} puede ser calculada por la siguiente ecuación:

$$\mathbf{T} = \mathbf{X}\mathbf{R} \quad (6)$$

Por consiguiente, el modelo final de regresión se puede expresar por:

$$\mathbf{y}_{\text{calc}} = \mathbf{X}\mathbf{R}\mathbf{b} = \mathbf{X}\beta_{\text{PLS}} \quad (7)$$

donde $\beta_{\text{PLS}} = \mathbf{R}\mathbf{b}$. Además, si se toman en cuenta todas las variables latentes, los coeficientes β_{PLS} son idénticos a los coeficientes del modelo de regresión lineal múltiple MLR (incluyendo todas las variables), cuando éste sea aplicable:

$$\beta_{\text{MLR}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (8)$$

Ventajas y desventajas del método PLS

El método obtiene, a partir de la matriz \mathbf{X} (variables independientes), una matriz \mathbf{T} (variables latentes) cuyos vectores son *linealmente independientes*, definiendo un sistema *ortogonal*. En consecuencia, se pueden usar los vectores \mathbf{t} de manera conveniente para hacer aproximaciones [19]. Además, en los casos en que existan un número grande de variables independientes ($m > n$, Figura 1) se produce una reducción del modelo. La información relevante es resumida en las primeras variables latentes, mientras que el "ruido" es modelado por las últimas. En los casos en que exista colinealidad (redundancia) entre las variables, ésta se usa para reducir el ruido aleatorio. Por consiguiente, es posible minimizar el riesgo de cometer un error estadístico de Tipo II (descartar información importante) [4]. Otro aspecto favorable del método es que produce una

reducción en la magnitud del vector β (Ecuación 1), una característica positiva para cualquier método de regresión [20]:

$$|\beta_{\text{PLS}, 1}| \leq |\beta_{\text{PLS}, 2}| \leq \dots \leq |\beta_{\text{MLR}}| \quad (9)$$

donde los subíndices se refieren a las variables latentes y β_{MLR} se define por la Ecuación 8. Una desventaja es que el modelo es correlativo y no causal, por consiguiente, los modelos obtenidos no ofrecen información fundamental acerca del fenómeno estudiado [20]. Una crítica más profunda, desde el punto de vista de la teoría estadística, es que PLS no puede ser considerado un método óptimo [20].

Ejemplo de Aplicación: Predicción de Propiedades de Carbones Minerales

El carbón es un recurso natural abundante en muchos países del mundo y, particularmente, en Venezuela. Extensos depósitos han sido reportados en los estados Zulia (Guasare, Socuy, Aricuaisá, Río de Oro, Casigua), Táchira (Santo Domingo, Rubio, Lobatera, Hato de la Virgen, Las Adjuntas, Franja Nororiental), Falcón (Pedregal) y Anzoátegui (Naricual, Fila Maestra), principalmente en las formaciones geológicas Marcelina, Los Cuervos, Carbonera, Cerro Pelado, Naricual y Quebradón [21]. Pese a su valor como fuente de componentes orgánicos, es empleado generalmente como combustible para la producción de calor y energía eléctrica, o como materia prima para producir coque, usado en siderurgia y metalurgia para la reducción de óxidos metálicos.

La caracterización del carbón mineral con fines del establecimiento de su calidad para las aplicaciones antes expuestas, comprende una serie de determinaciones analíticas de complejidad, costo y tiempo de respuesta variables. Los análisis petrográficos permiten reconocer el grado de evaluación del carbón mediante la medición de la reflectancia promedio de la vitrinita, según la norma ASTM D 2748-79 [22]. Los análisis inmediatos comprenden la determinación de los porcentajes de humedad (ASTM D 3173-73), materia volátil (D 3175-82), cenizas (D 3175-82) y carbono fijo (por diferencia al 100% de los tres anteriores). El azufre total se determina por el método de la bomba calorimétrica (D 3177-84).

El poder calórico se mide según la norma D 2015-85. Los análisis últimos o elementales incluyen las determinaciones de los tenores de C, H (tren de combustión; D 3178), N (Kjeldahl; D 3179); S (bomba calorimétrica; D 3177) y O (por diferencia) [22]. Una alternativa rápida, no destructiva, con poco trabajo de preparación de muestras y al alcance de cualquier laboratorio, esta dada por el uso de la espectroscopía de infrarrojo.

La espectroscopía de infrarrojo es una poderosa herramienta de la química analítica, que permite estudiar la estructura de materiales. Por medio de un espectro de infrarrojo, es posible conseguir información tanto de tipo cualitativo (tipo de grupos funcionales presentes en la muestra) como cuantitativo. A manera ilustrativa de la potencialidad de este método sobre carbones, se consideró un conjunto de 49 muestras de carbones venezolanos y colombianos, cuyas propiedades han sido publicadas [23-25]. Detalles del procedimiento experimental de obtención de los espectros se encuentran en el Trabajo Especial de Grado de C.E.A. [26]. Las propiedades: materia volátil, carbono fijo, cenizas, poder calórico, carbono elemental, hidrógeno, nitrógeno, azufre total, y reflectancia máxima de vitrinita ($R_o \max$) fueron determinadas por métodos convencionales (según normas ASTM) y correlacionadas con el espectro de infrarrojo. En este caso, la matriz \mathbf{X} está representada por los espectros de infrarrojo (en unidades de absorbancia) y el vector \mathbf{y} por los datos de referencia. Un espectro típico de infrarrojo se muestra en la Figura 3. Las asignaciones de bandas se encuentran en [26] y referencias allí incluidas. La absorbancia a cada uno de los números de onda considerados representa una variable independiente. Por consiguiente, en este caso la matriz \mathbf{X} representa 171 variables independientes ($m = 171$).

El número óptimo de variables latentes a utilizar se determinó por el método de "bootstrapping" mencionado anteriormente. El resultado para el porcentaje de hidrógeno se muestra en la Figura 4, donde el número óptimo de variables latentes es 6.

En la Tabla 3 se muestra el número de variables latentes y las desviaciones estándar de las predicciones, y se compara con las desviaciones estándar de los valores de referencia. Se obtuvo

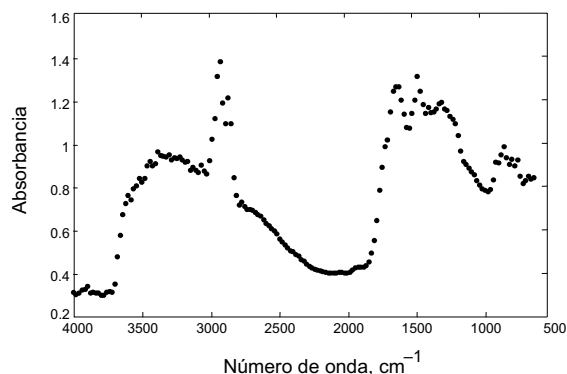


Figura 3. Un espectro típico de infrarrojo de una muestra de carbón mineral.

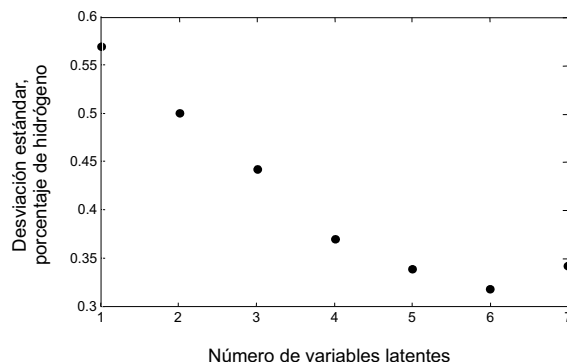


Figura 4. Desviaciones estándar vs. número de variables latentes para el porcentaje de hidrógeno predicho por PLS (validación cruzada).

buenas correlaciones para carbono fijo, cenizas, reflectancia de vitrinita y concentración de carbono e hidrógeno, demostrando la potencialidad de la espectroscopía de infrarrojo en el análisis rápido, económico y no destructivo del carbón mineral.

Notación

(Los vectores mostrados son vectores "columna". Los vectores "fila" se indican transpuestos, por ejemplo, \mathbf{w}^T).

- \mathbf{x} Variable independiente, ($n \times 1$).
- \mathbf{X} Matriz de variables independientes, ($n \times m$).
- \mathbf{y} Variable dependiente, ($n \times 1$).

Tabla 3
Predicciones de propiedades de carbones minerales usando el espectro de infrarrojo y PLS con validación cruzada (26)

Propiedad, rango y desviación estándar (D.E) de las muestras (ref. 26)	Número de variables latentes	SD (desviación estándar de las predicciones)
Materia volátil (% en peso, b.s.) 19.9-63.6; D.E. 11.1	5	3,5
Carbono fijo (% en peso, b.s.) 29.8-73.9; D.E. 11.7	8	4,6
Cenizas (% en peso, b.s.) 1.7-27.2; D.E. 5.3	2	3,8
Poder Calórico (MJ/kg, b.s.) 24.1-35.2; D.E. 2.4	3	1,9
Carbono elemental (% en peso, b.s.) 54.1-87.3; D.E. 6.3	9	4,1
Hidrógeno (% en peso, b.s.) 4.7-8.1; D.E. 0.96	6	0,32
Nitrógeno (% en peso, b.s.) 0.0-2.1; D.E. 0.49	1	0,49
Azufre total (% en peso, b.s.) 0.3-12.2; D.E. 2.6	6	1,6
Reflectancia máxima de vitrinita (Ro max) 0.34-1.36; D.E. 0.25	5	0,10

- β Vector de regresión, que multiplica a \mathbf{X} ($m \times 1$).
- \mathbf{E} Matriz residual de variables independientes, ($n \times m$).
- \mathbf{f} Vector residual (error) de la variable dependiente \mathbf{y} , ($n \times 1$).
- \mathbf{b} Vector de coeficientes de regresión, que multiplica a \mathbf{T} ($a \times 1$).
- \mathbf{w} Vector de "weights", ($m \times 1$).
- \mathbf{W} Matriz de "weights", ($m \times a$).
- \mathbf{t} Vector "scores", ($n \times 1$).
- \mathbf{T} Matriz de "scores", ($n \times a$).
- \mathbf{p} Vector de "loadings", ($m \times 1$).

Agradecimiento

Se agradece la asistencia financiera del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICIT) a través del Proyecto S1-2278.

Referencias Bibliográficas

1. Geladi P., Kowalski B. R.: "Partial Least Squares Regression: A Tutorial". Anal. Chim. Acta, Vol. 185, (1986) 1-17.
2. Geladi P., Kowalski B. R.: "An Example of 2-Block Predictive Partial-Least Squares Regression with Simulated data". Anal. Chim. Acta, Vol. 185, (1986) 19-32.
3. Wold S.: "Personal memories of the early PLS development". Chemom. Intell. Lab. Syst. Vol. 58 No. 2 (2001) 83-84.
4. Martens H.: "Reliable and relevant modeling of real world data: a personal account of the development of PLS regression". Chemom. Intell. Lab. Syst. Vol. 58 No. 2 (2001) 85-95.
5. Wold S., Trygg J., Berglund A., Antti H.: "Some recent developments in PLS modeling". Chemom. Intell. Lab. Syst. Vol. 58 No. 2 (2001) 131-150.

6. Lavine B. K.: "Chemometrics". *Anal. Chem.*, Vol. 70, No. 12 (1998) 209R-226R.
7. Sekulic S., Seasholtz M. B., Wang Z., Kowalski B. R., Lee S. E., Holt B. R.: "Nonlinear Multivariate Calibration Methods in Analytical Chemistry". *Anal. Chem.*, Vol. 65, No. 19 (1993) 835A-845A.
8. Erickson C. L., Lysaght M. J., Callis J. B.: "Relationship between Digital Filtering and Multivariate Regression in Quantitative Analysis". *Anal. Chem.*, Vol. 64, No. 24 (1992) 1155A-1163A.
9. Beebe K. R., Kowalski B. R.: "An Introduction to Multivariate Calibrations and Analysis". *Anal. Chem.*, Vol. 59, No. 17 (1987) 1007A-1017A.
10. Martens H., Naes T.: "Multivariate Calibration". John Wiley, Chichester 1989.
11. Mardia K. V.: "Multivariable Analysis". Academic Press, 1979.
12. Draper N., Smith H.: "Applied Regression Analysis". Wiley, New York 1981.
13. Qin, S. J.: "Recursive PLS algorithms for adaptive data modeling". *Computers Chem. Eng.* Vol. 22, No. 4-5 (1998) 503-514.
14. Wold, S., Ruke, A., Wold, H., & Dunn III, W.J.: "The collinearity problem in linear regression, the partial least squares (PLS) approach to generalized inverses". *SIAM Journal of Scientific and Statistical Computation.* Vol. 5, No. 3 (1984) 735-743.
15. Kaspar M. H., Ray W. H.: "Partial Least Squares Modelling as Successive Singular Value Decompositions". *Computers Chem. Eng.*, Vol. 17, No. 10 (1993) 985-989.
16. de Jong S., ter Braak C.J.F.: "Comments on the PLS kernel algorithm". *J. Chemometrics*, Vol. 8, (1994) 169-174.
17. de Jong S., Wise B.M., Ricker N.L.: "Canonical partial least squares and continuum power regression". *J. Chemometrics*, Vol. 15, No.2 (2001) 85-100.
18. Baffi G., Martín E. B., Morris A.J.: "Non-linear projection to latent structures revisited: the quadratic PLS algorithm." *Computers Chem. Eng.*, Vol. 23, (1999) 395-411.
19. Kincaid D., Cheney W.: "Análisis Numérico: Las Matemáticas del Cálculo Científico". Addison-Wesley Iberoamericana 1994.
20. Helland, I. S.: "Some theoretical aspects of partial least square regression". *Chemom. Intell. Lab. Syst.* Vol. 58 No. 2 (2001) 97-107.
21. Escobar M. E., Martínez M.: "Los depósitos de carbón en Venezuela". *Interciencia*, Vol. 18, No. 5 (1993) 224-229.
22. American Society for Testing Materials: "Gaseous fuels, coal and coke" in *Annual Book of ASTM Standards: Philadelphia, Section 5*, vol. 05.05, 1992
23. Alciaturi C. E., Escobar M. E., De La Cruz C., Rincón C.: "Predicción de propiedades de carbones minerales por espectroscopia infrarroja y regresión de componentes principales". *Memorias de las 10mas Jornadas Científico Técnicas de Ingeniería*, Vol. 2, (1998), QUI 40- QUI 45.
24. Alciaturi C. E., Montero T., De La Cruz C., Escobar M. E.: "The Prediction of Coal Properties using Compressed Infrared Data from Osculating Polynomials". *Anal. Chim. Acta*, Vol. 340, (1997), 233-240.
25. Alciaturi C. E., Escobar M. E., Vallejo, R.: "Prediction of Coal Properties by Derivative DRIFT Spectroscopy". *Fuel*, Vol. 34, No. 4 (1996), 491-499.
26. Alciaturi, Carlos E. Trabajo Especial de Grado, Universidad del Zulia, Facultad de Ingeniería, (2000).

Recibido el 10 de Marzo de 2003

En forma revisada el 8 de Diciembre de 2003