



# REVISTA MEXICANA DE FISICA

Revista Mexicana de Física  
Sociedad Mexicana de Física  
rmf@smf2.fciencias.unam.mx  
ISSN (Versión impresa): 0035-001X  
MÉXICO

2005

C. Araujo Andrade / I. Campos Cantón / J. R. Martínez / G. Ortega Zarzosa / F. Ruiz  
MODELO DE PREDICCIÓN BASADO EN ANÁLISIS MULTIVARIANTE PARA LA  
DETERMINACIÓN DE CONCENTRACIÓN DE AZÚCAR EN SOLUCIÓN

*Revista Mexicana de Física*, diciembre, año/vol. 51, número enseñanza 2

Sociedad Mexicana de Física  
Distrito Federal, México  
pp. 67-73

---

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

Universidad Autónoma del Estado de México

**redalyc**  
LA MEMORIA CIENTÍFICA EN LÍNEA  
<http://redalyc.uaemex.mx>

# Modelo de predicción basado en análisis multivariante para la determinación de concentración de azúcar en solución

C. Araujo-Andrade<sup>a,b</sup>, I. Campos-Cantón<sup>a</sup>, J.R. Martínez<sup>a</sup>, G. Ortega-Zarzosa<sup>a</sup> y F. Ruiz<sup>a</sup>

<sup>a</sup>Facultad de Ciencias, Universidad Autónoma de San Luis Potosí,  
78000 San Luis Potosí, S.L.P., México,

<sup>b</sup>Instituto Potosino de Investigación Científica y Tecnológica

Recibido el 10 de agosto de 2004; aceptado el 11 de octubre de 2004

La medición de la actividad óptica utilizando el principio de polarización es un método comúnmente usado para determinar la cantidad del contenido de azúcares en soluciones. El método de polarización es regularmente utilizado en la industria alimentaria. A fin de mostrar la importancia de los métodos estadísticos para el análisis de soluciones complejas, utilizamos un modelo de predicción basado en el análisis multivariante para la determinación del contenido de azúcar en soluciones medidas por técnicas espectroscópicas. Específicamente, medimos espectros en el cercano infrarrojo de soluciones con diferente concentración de azúcar, como base para la determinación de concentraciones desconocidas de azúcares. En particular, estimamos la concentración de azúcar en bebidas gaseosas comerciales, obteniendo resultados satisfactorios.

*Descriptores:* Álgebra lineal; espectroscopía de cercano infrarrojo; análisis multivariante; práctica de laboratorio.

Polarization principle is a rutinary method to measure the optical activity of solutions with sugar. In this work we use a prediction model based in multivariate analysis to show the applicability of statistical method in analysis of mixtures. We measure near infrared absorption of solution with different sugar concentration for the determination of unknown concentration of sugar. We apply the model to estimate the concentration of sugar in sodas.

*Keywords:* Linear algebra; near infrared spectroscopy; multivariate analysis; laboratory.

PACS: 01.50.Lc; 02.10.Ud; 33.20.Ea; 01.50.-i

## 1. Introducción

La polarimetría es una técnica que se basa en la medición de la rotación óptica producida sobre un haz de luz polarizada al pasar por una sustancia ópticamente activa. La actividad óptica rotatoria de una sustancia tiene su origen en la asimetría estructural de los átomos de carbono, nitrógeno, fósforo o azufre en la molécula, lo cual es conocido como quiralidad. La quiralidad generalmente es descrita como una imagen de espejo para una molécula, la cual no puede superponerse con ella misma.

El uso de polarímetros es una técnica extendida para la determinación de concentración de azúcar en soluciones. A través del polarímetro se mide la actividad óptica de soluciones, en la cual, luz con determinada longitud de onda es polarizada en un plano por un polarizador, y entonces estudiada por un segundo polarizador con el objeto de encontrar su nuevo plano de polarización después de haber pasado por la muestra en solución. La concentración de azúcar es proporcional al ángulo de polarización.

A fin de mostrar la importancia de los métodos estadísticos para el análisis de soluciones complejas, utilizamos un modelo de predicción basado en el análisis multivariante para la determinación del contenido de azúcar en soluciones medidas por técnicas espectroscópicas. Específicamente, medimos espectros en el cercano infrarrojo de soluciones con diferente concentración de azúcar, como base para la determinación de concentraciones desconocidas de azúcares. En particular, para mostrar la aplicabilidad del método estimamos el conte-

nido de azúcar en bebidas gaseosas comerciales, para lo cual medimos su respectivo espectro de absorción en el cercano infrarrojo y aplicamos el modelo de predicción desarrollado para cuantificar su contenido de azúcar.

## 2. Calibración multivariante

Las técnicas de calibración multivariante, tales como regresión parcial de mínimos cuadrados (PLS, por sus siglas en inglés) son capaces de entenderse con procesamiento de espectros y su cuantificación [1]. Una vez que la calibración ha sido realizada, la predicción de futuras muestras con concentraciones (determinación de concentraciones en muestras desconocidas por regresión) es meramente una multiplicación de matrices.

El método de regresión parcial de mínimos cuadrados es un método muy popular en aplicaciones industriales de modelación suave [2,3]. Investigaciones en ciencias e ingeniería comúnmente involucran el uso de variables controlables (factores) y/o fáciles de medir que explican, regulan, o predicen el comportamiento de otras variables (respuestas). Cuando los factores son pocos en número, no son significativamente redundantes (colineales) y tienen una bien entendida relación a las respuestas; de esta manera, una regresión lineal múltiple puede ser una buena manera de procesar datos de información. En aplicaciones de sistemas suaves el investigador se enfrenta a muchas variables y a relaciones complejas, de esta manera los objetos son usados para construir un buen modelo de predicción. El método PLS es un método para construc-

ción de modelos de predicción cuando los factores son muchos y altamente colineales. Nótese que el énfasis es sobre la predicción de las respuestas y no necesariamente en tratar de entender las relaciones existentes entre variables. El método PLS es una herramienta útil cuando el objetivo es la predicción y no existe una necesidad práctica de limitar el número de factores medidos.

El método PLS es un método de calibración multivariante, el cual es capaz de relacionar la concentración de elementos directamente con espectros medidos (o porciones de ellos), particularmente con NIR [4,5]. Las variables espectrales son colectadas como una matriz **X** con número de filas igual al número de muestras y el número de columnas igual al número de canales en el espectro.

La matriz **Y** consiste de las concentraciones (No. de filas = No. de muestras; No. de columnas = No. de elementos). La relación puede ser escrita como  $\mathbf{Y} = \mathbf{XB} + \mathbf{F}$ , en la cual la matriz **F** es la matriz que contiene los residuales (variaciones no descritas por el modelo). El coeficiente **B** puede ser calculado de diferentes maneras. Un método frecuentemente usado es el de regresión lineal múltiple (MLR), en el cual la solución de mínimos cuadrados es dada por  $\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ . Sin embargo, cuando el número de variables *x* excede el número de muestras o cuando existe un alto grado de colinealidad entre las variables, la solución resultante no es estable. Matemáticamente esto significa que la inversa de  $\mathbf{X}'\mathbf{X}$  no existe. El método PLS maneja este problema de colinearidad comprimiendo la matriz **X** de datos  $\mathbf{X} = [x_1, x_2, \dots, x_p]$  que contiene el espectro de *p* canales para las *n* muestras, en un número de *A* variables latentes ortogonales o *scores*  $\mathbf{T} = [t_1, t_2, \dots, t_A]$ . El número de variables latentes es usualmente mucho más pequeño que el número original de variables *p*. Estos *scores* **T** son usados para ajustar un conjunto de *n* observaciones a *m* variables dependientes de la concentración  $\mathbf{Y} = [y_1, y_2, \dots, y_m]$ . Como las variables latentes son ortogonales (linealmente independientes unas de otras) el inverso puede ser obtenido fácilmente, resolviendo así el problema de colinearidad [6].

El modelo PLS puede considerarse como consistente de dos relaciones externas y una relación interna. Existe una relación externa para la matriz **X**:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} = \sum_a^A \mathbf{t}_a \mathbf{p}'_a + \mathbf{E},$$

y para la matriz **Y**

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F} = \sum_a^A \mathbf{u}_a \mathbf{q}'_a + \mathbf{F}$$

En donde **P** y **Q** son los *loadings* del conjunto de variables **X** y **Y**, respectivamente, y **E** y **F** son matrices que contienen los residuales. *A* representa el número de variables latentes retenidas en el modelo PLS. Los *loadings* describen cómo las variables **X** y **Y** originales están relacionadas con los *scores* **T** y **U**. La relación interna es escrita como

$$u_a = b_a t_a.$$

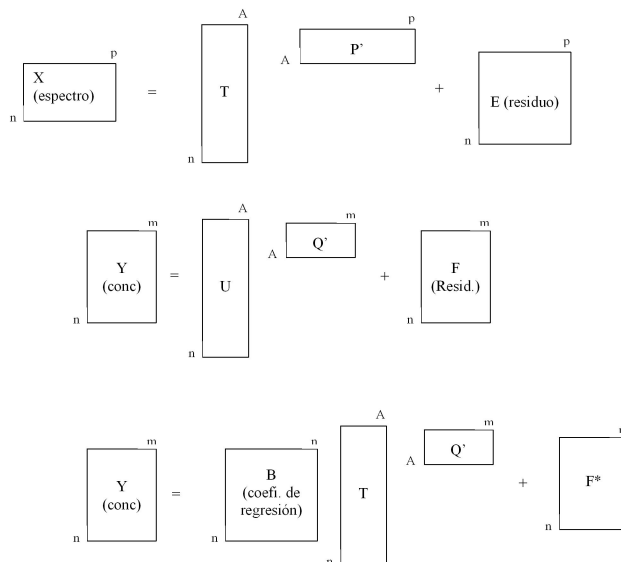


FIGURA 1. Representación gráfica de las relaciones internas y externas del modelo PLS.

En esencia, la relación interna es un ajuste de mínimos cuadrados entre el bloque de *scores* **X** y el bloque de *scores* **Y** con  $b_h = u'_h t_h / t'_h t_h$ . Cuando todos los *scores* y *loadings* son calculados, el modelo PLS final queda como

$$\mathbf{Y} = \mathbf{BTQ}' + \mathbf{F}^*.$$

La Fig. 1 muestra estas relaciones gráficamente.

Como mencionamos, el método PLS comprime las variables originales en un número de variables latentes. La primera variable latente primariamente tiende a describir la variancia sistemática en los datos (*i.e.*, la relación entre espectros y concentraciones) y en menor extensión el ruido. Como el número de *scores* se incrementa, la contribución del ruido se incrementa. Si éste es el caso, no se está modelando la variancia sistemática en los datos sino el ruido, entonces el modelo resultante tiene una pobre capacidad de predicción. El número óptimo de variables latentes o *scores* es por lo tanto muy importante. La derivación de la optimización es principalmente calculada de la raíz cuadrática media del error de predicción (RMSEP) para los diferentes números de variables latentes usadas en el modelo PLS. Este valor de RMSEP es una medida de lo bueno del ajuste entre las concentraciones conocidas y las predichas:

$$\mathbf{RMSEP} = \sqrt{\frac{\sum_{n=1}^N (Y_{pred,n} - Y_n)^2}{N}}.$$

Cuando los valores de RMSEP alcanzan un mínimo, se dice que el número *score* correspondiente es óptimo. Los cálculos del valor RMSEP son hechos comúnmente de dos formas: realizando una validación cruzada (LOOCV) [7],

usada en este trabajo, o a través de la separación de un conjunto de datos. La técnica LOOCV es generalmente utilizada cuando está disponible solamente un número limitado de muestras para construir y validar una calibración. Cada muestra es excluida del conjunto de datos y el modelo PLS es construido basado en la muestras remanentes.

Una vez construido el modelo de calibración, la muestra excluida es predicha y la desviación almacenada. Esto se realiza hasta que todas las muestras son excluidas una vez. El valor RMSEP es la media de las desviaciones almacenadas. En el caso del conjunto separado de datos, el modelo PLS es construido del llamado conjunto de entrenamiento y usado para predecir las muestras del llamado conjunto de predicción.

Además del valor RMSEP, el coeficiente de correlación ( $r$ ) es la medida de la relación lineal entre dos variables, el cual es un número absoluto sin unidades. Así cuando la covariancia,  $cov(x,y)$ , de dos variables se aproxima a  $-1$  ó  $1$ , se dice que existe una correlación lineal entre ellas. Los casos en donde la  $cov(x,y) = 1$ , representa una correlación perfecta positiva, y para la  $cov(x,y) = -1$ , representa una correlación perfecta negativa. El coeficiente de correlación refleja el grado de relación o efecto que tiene el cambio de una variable sobre la otra, la variancia  $r^2$  también es calculada. Su cuantificación nos indica cuánto de la variancia contenida en el conjunto de datos original es descrita por el modelo PLS. Esta cantidad se expresa como

$$r^2 = (S_{xy}S_{xy}/S_{xx}S_{yy}),$$

donde  $S_{xy}$  es la  $\sigma_{xy} = cov(x,y)$ , y cuando  $x = y$  la covarianza es igual a la variancia  $\sigma_x^2$ .

### 3. Método experimental

Para elaborar el modelo de predicción se prepararon catorce soluciones acuosas con concentración conocida de azúcar, disolviendo en agua destilada varias concentraciones de azúcar en un rango de  $0.02 \text{ g/cm}^3$  a  $0.36 \text{ g/cm}^3$ , para lo cual se disolvieron en 50 ml de agua destilada de 1g a 18 g de azúcar. La temperatura de las soluciones se mantuvo constante para todas las concentraciones. La solución se mantuvo a temperatura ambiente en un baño de temperatura constante durante 5 minutos previamente a su medición en el polarímetro y en el espectrofotómetro en el cercano infrarrojo.

A fin de comparar el modelo de predicción propuesto con los métodos estándares, se midieron los ángulos de polarización en términos de la concentración usando un polarímetro Phywe de media sombra, el cual utiliza la línea D del sodio con longitud de onda de 589 nm, y se midieron los espectros en el cercano infrarrojo usando un espectrofotómetro OceanOptics NIR512 miniatura de fibra óptica, con un detector de InGaAs y una lámpara de tungsteno-halógeno (w/900-hr, Bulbo de 3100 K, a 12 VDC).

El polarímetro se basa en la ley de Weber-Fechner, en la cual, a fin de incrementar la exactitud de las mediciones en

una mitad del campo de visión, existe una placa de cuarzo de Laurent la cual rota el plano de polarización a través de un ángulo pequeño. El analizador se ajusta hasta igualar la brillantez en los dos lados del campo de visión, lo que permite un ajuste fino y una fácil reproducción de las mediciones [8].

El procedimiento de preparación de las soluciones de azúcar en agua fue el siguiente: primeramente se calentó el agua en una estufa aproximadamente a  $50^\circ\text{C}$ , posteriormente se le agregaba el azúcar y se agitaba con un agitador magnético hasta que se disolviera todo el azúcar. Una vez realizado lo anterior, se obtuvieron los espectros de absorción en el cercano infrarrojo y en el polarímetro.

Para probar el modelo de predicción se seleccionaron una serie de bebidas gaseosas comerciales de las cuales se obtuvieron muestras que fueron medidas directamente en el espectrofotómetro NIR.

### 4. Resultados

Los ángulos de rotación de la luz polarizada causada por la solución de azúcar en agua en función de las concentraciones se presentan en la Fig. 2. En la figura se muestran los resultados experimentales y la curva ajustada mediante la línea continua, el comportamiento nos indica la dependencia lineal del ángulo de rotación en términos de la concentración.

Los espectros de absorción en el cercano infrarrojo se presentan en la Fig. 3 en el rango entre 850 nm y 1750 nm para las muestras con diferente concentración de azúcar. Como es de esperarse, observamos que los espectros son prácticamente los mismos y presentan muy pocas diferencias entre sí. Se pueden observar tres bandas de absorción localizadas en 960 nm, 1200 nm y 1500 nm. La banda centrada en 1500 nm es debida a vibraciones de grupos OH, la banda centrada en 1200 nm es asociada a vibraciones de grupos OH y CH y la banda centrada en 960 nm es asociada a grupos OH, esto de acuerdo con observaciones experimentales [9] y a lo reportado en la Ref. 10.

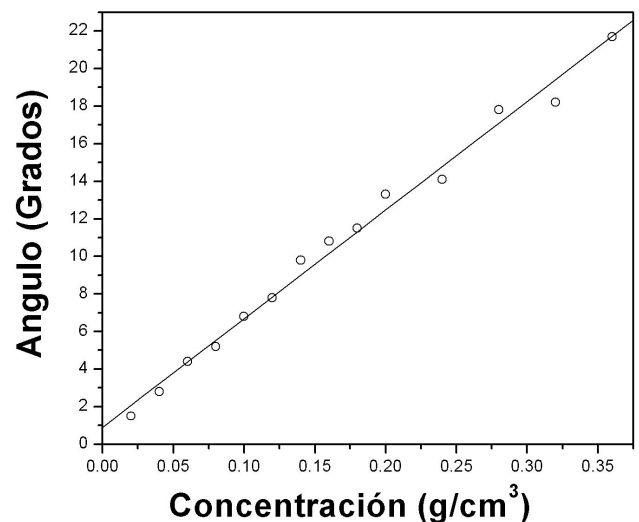


FIGURA 2. Ángulos de rotación de la luz polarizada para la solución de azúcar en agua en función de las concentraciones, la curva ajustada se muestra por la línea continua.

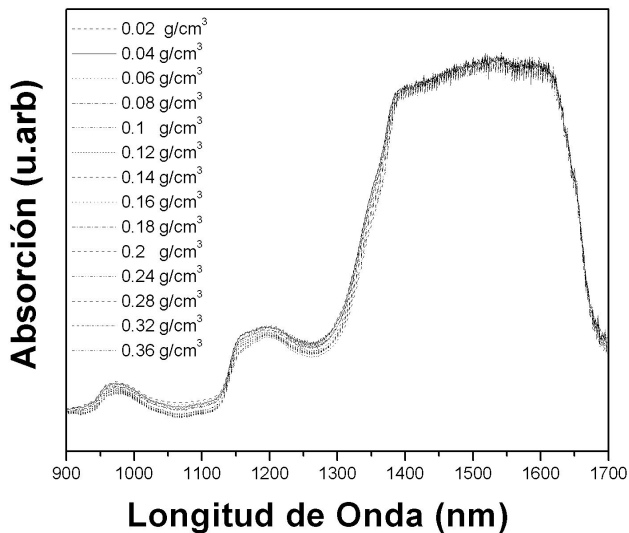


FIGURA 3. Espectros de absorción en el cercano infrarrojo para el rango entre 850 nm y 1750 nm de las muestras con diferente concentración de azúcar.

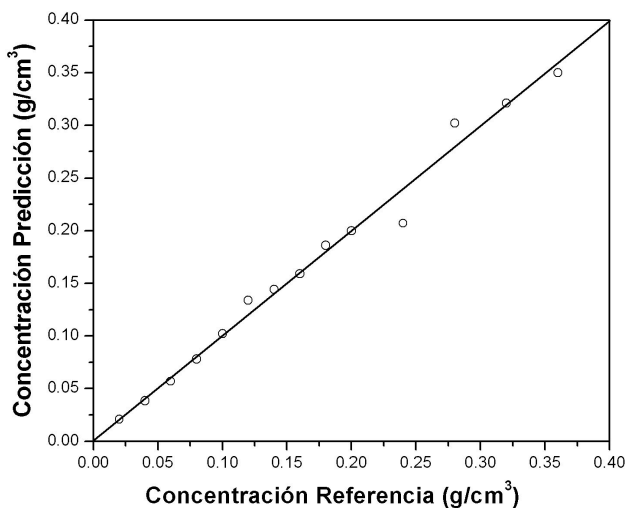


FIGURA 4. Validación cruzada para el modelo de predicción de concentración de azúcar ( $\text{g/cm}^3$ )

Como puede verse en la Fig. 3, es sumamente difícil observar cambios significativos entre las muestras, debido a la variación en el contenido de azúcar; prácticamente tenemos espectros sobrepuestos. La contribución de grupos CH en el espectro de absorción, que es la que se espera refleje los mayores cambios en el espectro debido a la presencia de azúcar en la solución, es oscurecida por la banda más intensa, correspondiente a grupos OH que se presenta en el mismo rango de longitud de onda. En particular, la contribución de grupos CH se manifiesta en los 1180 nm y los cambios asociados en esta región, debido a la variación de concentración de azúcar, es imperceptible y así los resultados NIR por sí solos no son de utilidad. En este punto es que la utilización de métodos estadísticos se hace importante. En particular, a los resultados de absorción en el cercano infrarrojo les aplicaremos la técnica de calibración multivariante PLS para poder cuantificar

los pequeños cambios en el espectro de absorción debidos a la variación de concentración de azúcar.

Los espectros de absorción NIR, fueron analizados utilizando la técnica de calibración multivariante PLS (*Partial least square regression*). Para su análisis, existen programas comerciales que permiten realizar la calibración multivariante, como el programa de computo Unscrambler [11], el cual trabaja con base en estos métodos de análisis multivariante, para construir un modelo de predicción para las soluciones azúcar/agua; o bien puede utilizarse el software Matlab (The Math Works Inc.) [12] para escribir un programa que incluye procedimientos recursivos para la resolución de la matriz de datos, como el utilizado en este trabajo.

La validación cruzada para el modelo de predicción de concentración de azúcar ( $\text{g/cm}^3$ ), se muestra en la Fig. 4. Los resultados numéricos se muestran en la Tabla I, para estos resultados se obtuvo un coeficiente de correlación de 0.993, y un error RMSEP de  $0.012 \text{ g/cm}^3$ .

El programa PLS empleado es un programa escrito en el laboratorio que incluye procedimientos recursivos para la resolución de la matriz de datos y cuyo algoritmo se presenta en la Tabla II. Este programa fue escrito para determinar la agrupación de la evolución y su comparación con otras mezclas. El rango de longitud de onda gravada por el espectro fue de 850 nm a 1750 nm. Dicha región se empleó para su análisis multivariante, con la técnica de PLS, siendo la matriz de datos original de  $8 \times 512$ , donde cada fila corresponde a una muestra y cada columna a una longitud de onda, los datos fueron preprocesados utilizando un método de autoescalado.

Los espectros NIR para las distintas concentraciones de azúcar en agua fueron etiquetados con su concentración respectiva, obteniéndose una matriz de espectros y una matriz de concentraciones, como se ejemplifica en las Tablas III

TABLA I. Resultados numéricos de la validación cruzada para el modelo de predicción de concentración de azúcar.

Concentración de Referencia ( $\text{gr/cm}^3$ )	Concentración predicha ( $\text{gr/cm}^3$ )
0.02	0.02101
0.04	0.03841
0.06	0.05708
0.08	0.07809
0.10	0.102
0.12	0.134
0.14	0.144
0.16	0.159
0.18	0.186
0.20	0.200
0.24	0.207
0.28	0.302
0.32	0.321
0.36	0.350

TABLE II. Algoritmo Programa PLS escrito en el laboratorio que incluye procedimientos recursivos para la resolución de la matriz de datos.

---

```

clear
[Spectra, n] = XLSREAD('ESP01');
[Concs, n] = XLSREAD('py01');
LoadNum = 10;

% coloca a cero las variables globales
global w b v MeanA MeanC tt;
w=[]; b=[]; v=[]; MeanA=[]; MeanC=[]; tt=[];

%paso 1: preprocesado
MeanA=mean(Spectra);
A = Spectra- ones(size(Spectra,1), 1)*MeanA;

MeanC=mean(Concs);
c = Concs-ones(size(Spectra,1), 1)*MeanC;

h=1;

%lazo
while (h<=LoadNum)

    %paso 2: forma el peso del vector y normalizalo
    x=(A'*c)/(c'*c);
    w=[w,x/sqrt(x'*x)];

    %paso 3: forma el vector de scores latentes
    t=A*w(:,h);
    %calcula la suma de los cuadrados, ya que posteriormente son usados
    tt=t'*t;

    %paso 4: relaciona el vector de scores con las concentraciones
    v=[v;(t'*c)/tt];

    %paso 5:forma B, el vector de cargas PLS
    b=[b,((A'*t)/tt)];

    %paso 6: calcula los residuos
    A=A-t*b(:,h)';
    % c=c-v(h)*t;
    c=c-t*v(h,:);

    %paso 7: Incrementa y cierra el lazo
    h=h+1;

end;

%estimación de las concentraciones
best = w*(inv(b'*w)*v);
bo = MeanC - MeanA*best;
yest = ones(size(Spectra,1), 1)*bo + Spectra*best;

```

---

TABLA III. Representación formal de la matriz de concentraciones ( $Y_r$ ) y la matriz de espectros (X).

$Y_r$	$X_{n,s}$									
$y_1$	$x_{1,1}$	$x_{2,1}$	$x_{3,1}$	$x_{4,1}$	$x_{5,1}$	$x_{6,1}$	$x_{7,1}$	$x_{8,1}$	...	$x_{n,1}$
$y_2$	$x_{1,2}$	$x_{2,2}$	$x_{3,2}$	$x_{4,2}$	$x_{5,2}$	$x_{6,2}$	$x_{7,2}$	$x_{8,2}$	...	$x_{n,2}$
$y_3$	$x_{1,3}$	$x_{2,3}$	$x_{3,3}$	$x_{4,3}$	$x_{5,3}$	$x_{6,3}$	$x_{7,3}$	$x_{8,3}$	...	$x_{n,3}$
$y_4$	$x_{1,4}$	$x_{2,4}$	$x_{3,4}$	$x_{4,4}$	$x_{5,4}$	$x_{6,4}$	$x_{7,4}$	$x_{8,4}$	...	$x_{n,4}$
$y_5$	$x_{1,5}$	$x_{2,5}$	$x_{3,5}$	$x_{4,5}$	$x_{5,5}$	$x_{6,5}$	$x_{7,5}$	$x_{8,5}$	...	$x_{n,5}$
$y_6$	$x_{1,6}$	$x_{2,6}$	$x_{3,6}$	$x_{4,6}$	$x_{5,6}$	$x_{6,6}$	$x_{7,6}$	$x_{8,6}$	...	$x_{n,6}$
$y_7$	$x_{1,7}$	$x_{2,7}$	$x_{3,7}$	$x_{4,7}$	$x_{5,7}$	$x_{6,7}$	$x_{7,7}$	$x_{8,7}$	...	$x_{n,7}$
$y_8$	$x_{1,8}$	$x_{2,8}$	$x_{3,8}$	$x_{4,8}$	$x_{5,8}$	$x_{6,8}$	$x_{7,8}$	$x_{8,8}$	...	$x_{n,8}$
$y_9$	$x_{1,9}$	$x_{2,9}$	$x_{3,9}$	$x_{4,9}$	$x_{5,9}$	$x_{6,9}$	$x_{7,9}$	$x_{8,9}$	...	$x_{n,9}$
...	...	...	...	...	...	...	...	...	...	...
$y_s$	$x_{1,s}$	$x_{2,s}$	$x_{3,s}$	$x_{4,s}$	$x_{5,s}$	$x_{6,s}$	$x_{7,s}$	$x_{8,s}$	...	$x_{n,s}$

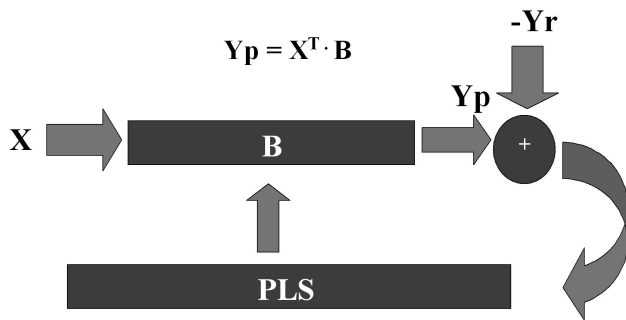


FIGURA 5. Diagrama que ejemplifica los pasos para generar el modelo de predicción.

y IV donde se representan de manera formal y numérica; en las tablas,  $Y_r$  es la matriz de concentraciones y X representa la matriz de espectros.

TABLA IV. Representación numérica de la matriz de concentraciones ( $Y_r$ ) y la matriz de espectros (X).

$Y_r$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	...	$X_n$
5	0.4727	0.4737	0.4747	0.4767	0.4777	0.4787	0.4807	0.4807	...	0.4727
10	0.4737	0.4737	0.4747	0.4757	0.4777	0.4787	0.4797	0.4797	...	0.4737
15	0.4739	0.4739	0.4759	0.4769	0.4799	0.4799	0.4819	0.4819	...	0.4739
20	0.4729	0.4739	0.4749	0.4759	0.4779	0.4779	0.4799	0.4809	...	0.4729
25	0.4727	0.4737	0.4747	0.4767	0.4777	0.4777	0.4797	0.4797	...	0.4727
30	0.4747	0.4737	0.4757	0.4747	0.4767	0.4767	0.4807	0.4807	...	0.4747
35	0.4727	0.4737	0.4757	0.4767	0.4787	0.4777	0.4797	0.4797	...	0.4727
40	0.4740	0.4740	0.4760	0.4750	0.4770	0.4780	0.4800	0.4810	...	0.4740
45	0.4737	0.4737	0.4757	0.4767	0.4787	0.4797	0.4817	0.4827	...	0.4737
...	...	...	...	...	...	...	...	...	...	...
S	0.4729	0.4739	0.4749	0.4759	0.4779	0.4779	0.4799	0.4809	...	0.4727

PREDICCIÓN

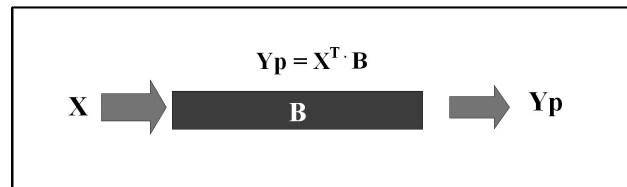


FIGURA 6. Diagrama que muestra la predicción, multiplicación vectorial de la matriz de espectros X y el vector de regresión para obtener la concentración escalar para cada espectro.

En la Fig. 5 se muestra el diagrama que ejemplifica los pasos para generar el modelo de predicción, para lo cual es necesario generar un vector de regresión, comúnmente llamado **b-vector**, el cual es generado con la técnica de mínimos cuadrados parciales (PLS), utilizando la matriz de concentraciones ( $Y_r$ ) y la matriz de espectros (X).

Una vez generado el b-vector ya no es necesario conocer la referencia ( $Y_r$ ), es decir solo con la matriz de espectros es posible obtener la concentración escalar para cada espectro, mediante la multiplicación vectorial de la matriz de espectros X y el vector de regresión, como se muestra en el diagrama de la Fig. 6.

Con el modelo de predicción construido es suficiente medir un espectro NIR de una solución cualquiera con contenido desconocido de azúcar, procesarla con el modelo de predicción y obtener la concentración de azúcar de dicha solución.

En particular, una de las concentraciones presentadas en la Fig. 3 y Tabla I era desconocida para el modelo, misma que fue reproducida satisfactoriamente de acuerdo a los valores del RMSEP. Para mostrar la aplicabilidad de la técnica, así como del modelo de predicción desarrollado, se presentan resultados de bebidas gaseosas comerciales que fueron medidas por espectroscopía en el cercano infrarrojo y cuyos datos fueron procesados en el modelo de predicción. En la Tabla V se muestran las bebidas seleccionadas así como su valor de concentración de azúcar predicho comparado con el valor reportado en la etiqueta. En la Fig. 7 se muestran los espectros de absorción NIR medidos y procesados en el modelo de predicción.

TABLE V. Resultados numéricos de la medición de concentración de azúcar de bebidas comerciales, mediante el modelo de predicción.

Muestra	Concentración medida (gr/ml)	Concentración reportada en etiqueta (gr/ml)
Coca Cola	0.161	0.110
Coca Cola Dieta	$2.5 \times 10^{-2}$	0.0
Fanta	0.188	0.118
Fresca	0.287	0.111

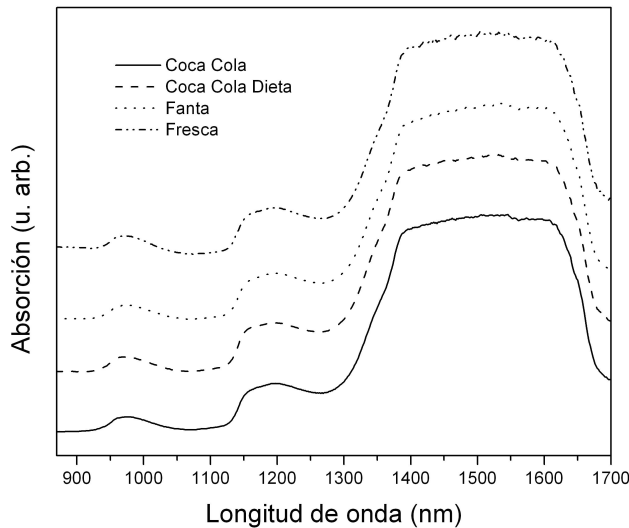


FIGURA 7. Espectros de absorción en el cercano infrarrojo para el rango entre 850 nm y 1750 nm de las muestras de bebidas comerciales.

### 5. Conclusiones

Las técnicas espectroscópicas, y en particular la de absorción en el cercano infrarrojo han demostrado ser muy eficientes para la detección de compuestos orgánicos en solución, debido a que en el cercano infrarrojo los efectos de fluorescencia son minimizados. Así, combinadas con las técnicas de quimiometría, es posible cuantificar y monitorear *in situ*, de una forma rápida y sencilla, la concentración de dicho compuesto. En este trabajo se usaron ambas técnicas para cuantificar la concentración de azúcar en solución acuosa, comparando con el método común de medición de concentración de azúcar por polarimetría, obteniendo resultados equivalentes. Las concentraciones de azúcar para bebidas comerciales, obtenidas a través de la técnica y del modelo presentado, arrojan resultados satisfactorios y equivalentes a los reportados en la etiqueta del producto.

1. H. Martens y T. Naes, *Multivariate Calibration* (Wiley, Chichester, 1991).
2. P. Lemberge y J. Van Espen, *X-Ray Spectrom.* **28** (1999) 77.
3. C. Demir y R.G. Brereton, *Analyst* **123** (1998) 181.
4. M.P. Fuller, G.L. Ritter y C.S. Draper, *Appl. Spectrosc.* **42** (1988) 2.
5. M.B. Seasholtz, D.D. Archibald, A. Lorber y B.R. Kowalski, *Appl. Spectrosc.* **43** (1989) 6.
6. R.G. Brereton, *Analyst* **125** (2000) 2125.
7. I.E. Frank y R. Todeschini, *The Data Analysis Handbook* (Elsevier, Amsterdam, 1994).
8. Phywe series of publications, *Laboratory Experiments* (LEP 2.5.02, Göttingen, Germany).
9. G. Yoon, A.K. Amerov, K.J. Jeon y Yoen-Joo Kim, *Appl. Opt.* **41** (2002) 1469.
10. D.K Pedersen y S.B.Engelsen, *New Food* (2002) 9.
11. The Unscrambler v. 8.0, CAMO Process As. (1986-2003).
12. Matlab: *The Language of Technical Computing*, The MathWorks, Inc. (1999).