

Uso del análisis multivariado de datos y espectrometría de absorción Uv-visible para la determinación simultánea de la concentración de Co, Cu y Ni en una solución: Reporte preliminar

Germán Comina⁽¹⁾ gcomina@uni.edu.pe; José Solís^(2,1) jsolis@ipen.gob.pe;
Walter Estrada^(2,1) westrada@ipen.gob.pe

(1) *Universidad Nacional de Ingeniería. Facultad de Ciencias, Av. Tupac Amaru 210, Lima 25, Perú*

(2) *Instituto Peruano de Energía Nuclear (IPEN), Av. Canadá 1470, Lima 41, Perú*

Resumen

Se ha determinado en forma simultánea la concentración de Co, Cu y Ni en soluciones con 8% de error promedio. Para lograrlo, se desarrolló un modelo lineal usando el método PLS (Partial Least Square Regression). Además, se usó el análisis de componentes principales (PCA, Principal Components Analysis) para clasificar a las muestras por su densidad óptica total en el rango de medida, encontrándose que la densidad óptica se encuentra distribuida sobre la primera componente principal.

1. Introducción

La espectrofotometría UV-visible, combinada con la ley de Lambert-Beer, permite determinar la concentración de un soluto en una solución [1]. Una curva de calibración se obtiene usando soluciones de concentración conocida, y la absorbancia se mide a una longitud de onda donde se tiene la banda de absorción correspondiente al complejo que esta relacionada con el elemento de análisis. Con éste método se puede determinar la concentración de un soluto en la solución. Si se desea determinar la concentración de otros solutos en la misma solución, se tiene que repetir el mismo proceso para cada uno de ellos. En el presente trabajo utilizando análisis multivariado se determina en forma simultánea, luego del desarrollo de un modelo lineal, la concentración de Co, Cu, y Ni en una solución.

El análisis multivariado consiste en observar o medir ciertas propiedades de un sistema a través de mediciones indirectas de las mismas.

Las aplicaciones del análisis de datos usando métodos multivariados son bastante amplias, como por ejemplo en análisis de olor y sabor en narices y lenguas electrónicas [2, 3], en la determinación de variables de control ambiental, como la demanda química de oxígeno (COD) [4], el reconocimiento automático de rostros [5], etc.

Los objetivos del análisis multivariado se pueden dividir en tres grupos [6]:

- Descripción de datos (modelos para la exploración de datos).
- Clasificación y discriminación (PCA).
- Predicción y regresión (PLS, PCR).

El análisis por componentes principales (PCA), es la técnica más utilizada del análisis multivariado [7]. Su objetivo es identificar las fuentes que producen la mayor variancia en un conjunto de datos. Las variancias máximas del sistema se encuentran distribuidas en las llamadas componentes principales.

La medición de j parámetros o "variables" de i muestras u "objetos", pueden representarse por la matriz: X_{ij} . Por tanto cada medición de una muestra, se expresa como un vector en el espacio j -dimensional. El PCA permite reducir las dimensiones de dicho espacio.

Usando PCA los datos del espacio j -dimensional, por lo general, se proyectan en el plano formado por la componente principal 1 (PC1, es la dirección en el espacio j -dimensional, donde está la mayor variancia de X) y la componente principal 2 (PC2, es la segunda dirección en importancia donde esta la variancia de X); pero también es posible proyectar los puntos sobre el volumen formado por las primeras tres componentes principales.

El PCA usa el algoritmo NIPALS (Non-linear Iterative Projections by Alternating Least-Squares), inventado por Herman Wold en 1966. El primer paso es centrar cada objeto de la matriz respecto de la media \bar{x}_j como se observa en la ecuación:

$$x_{ij} = x_{ij} - \bar{x}_j \quad (1)$$

Luego se crea un modelo lineal, el cual consta por un término llamado de estructura, formado por el producto de dos matrices (TP^T), y el término de error residual, E , o la parte de la matriz que no pudo ajustarse al término de estructura, como se observa en la ecuación:

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

O en forma equivalente:

$$X = \sum_{w=1}^A t_w p_w^T + E \quad (3)$$

Donde A , es el número de componentes principales del modelo, t_w es el vector de "score" para la componente w y p_w es el vector de "loading" para la componente w . El algoritmo NIPALS calcula sucesivamente cada t_w , p_w y cada E_w que es la matriz de error para cada componente principal w . Los gráficos que se utilizan en el presente trabajo son los gráficos de "score" con dos componentes principales PC1 y PC2.

El método de regresión por mínimos cuadrados parciales (PLS, Partial Least-Square), ha ganado gran importancia en los campos de la química, física, medicina, economía y control de procesos industriales. El pionero en el uso de PLS en el campo de la economía fue Herman Wold a fines de los años sesenta. Los primeros en usar el método de PLS en aplicaciones químicas fueron S. Wold y H. Martens a fines de los años setenta luego de una aplicación hecha por Kowalski [8].

El método PLS crea un modelo lineal a partir del cual es posible predecir el valor de una o varias variables de un sistema. El método utiliza dos matrices X_{ij} y Y_{ik} . La matriz X , al igual que en el caso del PCA, representa a i objetos de j variables independientes. La matriz Y representa a i objetos o mediciones de k variables, para las cuales se desea construir un modelo, de tal manera se puede predecir los valores de las k variables, en mediciones futuras (para otras matrices X).

El proceso de calibración y predicción se observa en la Figura 1, el primer paso es la calibración, en este paso se establece un modelo a partir de las matrices conocidas X e Y . Luego teniendo el modelo, el segundo paso es usar el modelo para predecir valores de Y' a partir de los datos X' .

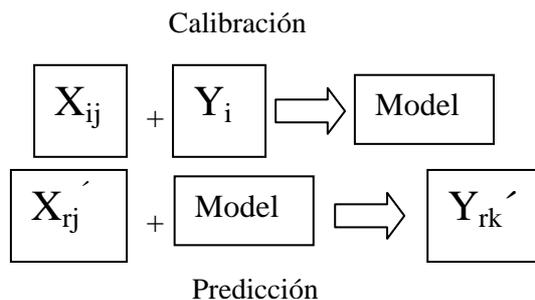


Figura 1. Procesos de calibración y predicción en el método PLS.

El modelo del PLS se basa en el algoritmo NIPALS. Se centran y redimensionan las matrices X e Y .

$$x_{ij} = x_{ij} - \bar{x}_j \quad (4)$$

$$y_{ik} = y_{ik} - \bar{y}_k \quad (5)$$

Luego se expresan independientemente como una suma de matrices, como en el caso del PCA.

$$X = \sum_{w=1}^A t_w p_w^T + E \quad (6)$$

$$Y = \sum_{w=1}^A u_w q_w^T + F^* \quad (7)$$

Luego los scores de cada uno de los modelos se relacionan en forma lineal, encontrando los llamados coeficientes de regresión, b_s :

$$b_s = \frac{u_s^* t_s}{t_s^* t_s} \quad (8)$$

La matriz Y puede entonces expresarse como:

$$Y = \sum_{w=1}^A t_w b_w q_w^* + F \quad (9)$$

Donde la condición para que el modelo sea aceptable es que $\|F\|$ sea lo suficientemente pequeño. Una medida utilizada para la aceptación de un modelo, es el error estándar de validación (SEV, Standard Error of Validation) el cual debe ser pequeño.

En el presente trabajo se determinaron las concentraciones de Co, Cu y Ni en tres soluciones problema diferentes, a partir de las medidas de sus espectros de absorción UV-visible. Se utilizaron 29 muestras de calibración, las cuales contenían Co, Cu, Ni, con concentraciones conocidas. Se usó el método PCA, con el cual se encontró una relación entre la primera componente principal y la densidad óptica espectral. Utilizando el análisis PLS de datos

multivariados se logro desarrollar un modelo con el cual se podrá predecir las concentración de Co, Cu y Ni en muestras.

2. Procedimiento Experimental

Se fabricaron tres soluciones base de Co, Cu, y Ni con una concentración de 1M, a partir de $\text{CoSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}$ y $\text{Ni}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$. Se obtuvieron 29 soluciones de calibración con diferentes concentraciones de Co, Cu, Ni en cada una. En la Tabla 1, se muestran las concentraciones de Co, Cu y Ni presentes en las muestras de calibración.

Tabla 1. Concentración de Co, Cu, y Ni presentes en las muestras de calibración.

Muestra	Co (mM)	Cu (mM)	Ni (mM)	SUMA (mM)
1	30	10	15	56
2	35	15	15	67
3	20	35	25	83
4	25	5	5	39
5	5	5	5	20
6	5	35	45	91
7	30	50	35	122
8	25	15	45	93
9	50	25	5	89
10	20	5	50	85
11	15	5	40	71
12	50	10	30	102
13	5	25	50	93
14	15	40	10	79
15	15	25	30	85
16	45	45	30	136
17	25	50	25	117
18	45	35	40	138
19	25	15	10	69
20	10	20	10	60
21	45	35	10	111
22	25	50	5	102
23	10	10	45	88
24	30	20	35	109
25	25	25	50	125
26	40	25	10	101
27	20	50	30	127
28	50	50	50	178
29	10	10	5	25

Además se obtuvieron tres soluciones problema, las cuales poseen diferentes concentraciones de Co, Cu y Ni; como se observa en la Tabla 2.

Se midió la absorbancia espectral de todas las soluciones utilizando un

espectrofotómetro UV-Visible de absorción Hitachi U-2000, en el rango de la longitud de onda de 350 nm a 800 nm; en pasos de 10 nm.

3. Resultados

Las mediciones se realizaron en forma aleatoria y para las muestras problema, se midió tres veces y en forma aleatoria, cada espectro.

Tabla 2. Concentración de Co, Cu, y Ni presentes en las muestras problema.

Muestra	Co (mM)	Cu (mM)	Ni (mM)
MP1	34	13	17
MP2	15	7	5
MP3	0	40	10

En la Figura 2 se muestran los 29 espectros de absorción para las muestras de calibración. Se tiene que cada pico de absorción corresponde a cada metal, los cuales están alrededor de 400, 500 y 800 nm, y corresponden a Ni, Co y Cu, respectivamente.

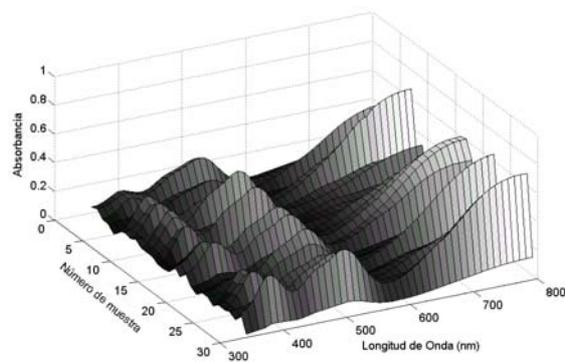


Figura 2. Espectros de absorbancia para las 29 muestras de calibración.

Los espectros de absorción para las tres muestras problemas se muestran en la Figura 3.

Observando la Figura 3 es posible ver que la forma de las curvas para las soluciones problema 1 y 2 son similares a las que se encuentran en la Figura 2, de tal manera que es posible decir que aquellas tienen disueltas los 3 metales, por que todas las soluciones de calibración tienen al menos 5 mM de cada metal. Para la solución problema 3, no presenta el pico de absorción que esta

alrededor de 500 nm. Por lo que se puede inferir que el Co no está presente en la solución. La absorbancia de cada muestra se midió 3 veces en diferente orden.

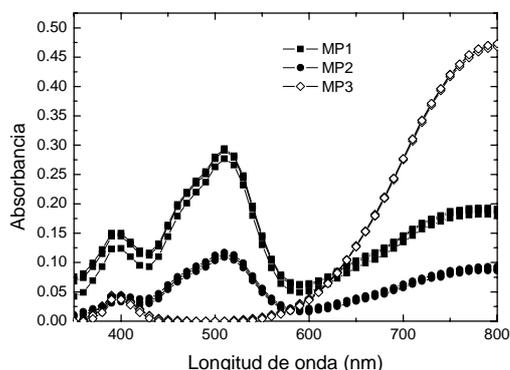


Figura 3. Absorbancia vs. longitud de onda para las 3 muestras problema.

Se utilizó el algoritmo NIPALS para determinar los scores y la gráfica PCI vs PC2 del método PCA. En la Figura 4 se muestra el gráfico de los Scores en las dos primeras componentes principales. Se observa el valor de la concentración total de las muestras de calibración, ver en la Tabla 1 la columna SUMA, están a la derecha del origen mientras que las que tienen menor concentración están a la izquierda en el gráfico PCA (Figura 4).

Por tanto, el valor de la suma de las concentraciones totales de las muestras (ver Tabla 1) están en relación directa con la primera componente principal del PCA. La concentración total va de mayor a menor cuando en el eje PC1 se va de izquierda a derecha.. O lo que es equivalente la concentración total está contenida en la primera componente.

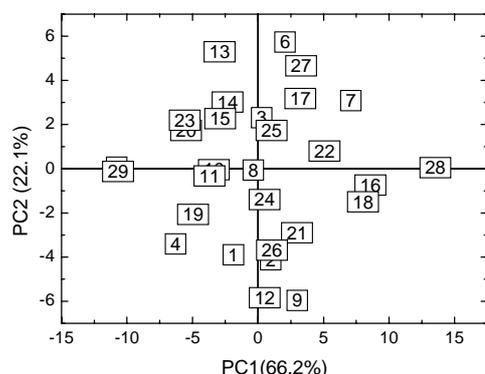


Figura 4. Gráfico PCA para las 29 muestras patrónes.

La ley combinada de Lambert-Bear [9], indica que hay una proporción directa entre el valor de la concentración de un soluto en

una solución y el valor de su densidad óptica (Densidad Óptica = $\log(1/T)$, T: transmitancia).

Por tanto la densidad óptica total medida en el rango de medida está contenida en la primera componente principal PC1. (Figura 4 y Tabla 1).

4. Conclusiones

Usando el método PCA, se observó que para el sistema en estudio la densidad óptica total evaluada en todo el rango de medida se encuentra sobre la primera componente principal.

Bibliografía

1. R.A.Day, A.L Underwood. Química Analítica Cuantitativa, 5ª Ed. Prentice-Hall (1989).
2. E.L. Kalman, F. Winqvist and I. Lundström, Atmospheric Environment, Vol.31, Issue 11, pp.1715-1719,(1997).
3. F. Winqvist, S. Holmin, C. Krantz-Rülcker, P. Wide and I. Lundström, Analytica Chimica Acta Vol.406, Issue 2, pp 147-157,(2000).
4. G. Comina, S. Holmin, P. Ivarsson, F. Winqvist, C. Krantz-Rülcker, 2nd. SENSPOL Workshop: Response to New Pollution Challenges. ISBN: 1 871315 86 7,p.22,.Londres, UK; 4-7 Junio 2002.
5. Vytautas Perlibakas, Patten Recognition Letters, Vol.25 Issue 6 pp.711-724 (2004).
6. K. Esbensen. Multivariate Data Analysis in Practice. CAMO Process AS, (2001).
7. Karl S. Booksh: Encyclopedia of Analytical Chemistry, R.A. Meyers (Ed.), John Wiley & Sons Ltd, pp. 8145-8169, Chichester, (2000).
8. Paul Geladi and Bruce R.Kowalski, Analytica Chimica Acta, Vol.185, pp.1-17, (1986).
9. Douglas Skoog, Principles of instrumental analysis, Brooks Cole 5th edition (1997).