

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

UNIVERSIDAD NACIONAL DE LA PAMPA

TRABAJO DE TESIS DE GRADO

Para optar el grado de

LICENCIADO EN QUÍMICA

**DETERMINACIÓN SIMULTÁNEA DE SACARINA Y ASPARTAMO EN
EDULCORANTES COMERCIALES UTILIZANDO HERRAMIENTAS DE
CALIBRACIÓN MULTIVARIADA**

Presentado por

MARIANELA SAVIO

Director

Dr. JOSÉ MANUEL CAMIÑA

Co-director

MSc. MARÍA SUSANA BOERIS

SANTA ROSA (LA PAMPA). ARGENTINA

2009

PREFACIO

Esta Tesina es presentada como parte de los requisitos para optar al grado Académico de Licenciado en Química de la Universidad Nacional de La Pampa y no ha sido presentada previamente para la obtención de otro título en esta Universidad ni en otra Institución Académica. Se llevó a cabo en el Departamento de Química de la Facultad de Ciencias Exactas y Naturales, durante el período comprendido entre el 27 de junio de 2008 (Resolución N° 192/08) y el 27 de Febrero de 2009 (Resolución N° 44) bajo la dirección del Dr. José Manuel CAMIÑA y bajo la codirección de la MSc. María Susana BOERIS.

Agradezco a mi Director de Tesis, Dr. José Manuel Camiña; a mi Co-Directora, MSc. María Susana Boeris; y al Lic. Miguel Cantarelli; por el acompañamiento y la ayuda brindada en el desarrollo de la Tesina y a lo largo de la carrera.

De igual manera agradezco a todos los docentes del Departamento de Química de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa, por todos los conocimientos brindados en estos años.

A todo el personal no docente por la excelente predisposición durante el desarrollo de la carrera.

Además agradezco de forma muy especial a mi Familia y a Mauro, por acompañarme y apoyarme en todo momento.

También agradezco a mis compañeros de carrera, especialmente a Antonela Iturri por haber compartido conmigo estos años.

Finalmente, deseo agradecer a todas las personas que de una u otra manera han sido partícipes con su colaboración y consejos en el presente trabajo.

Marianela SAVIO

Departamento de Química

Facultad de Ciencias Exactas y Naturales

Universidad Nacional de La Pampa

RESUMEN

En este trabajo, es propuesto un método nuevo, rápido y de bajo costo para la determinación simultánea de dos edulcorantes artificiales, aspartamo y sacarina. El modelo de calibración fue obtenido utilizando un diseño de experimento factorial completo, utilizando calibración multivariada en su variante PLSR-2 (Regresión por Mínimos Cuadrados Parciales). Los datos instrumentales fueron obtenidos por espectrofotometría UV-Visible. Por otra parte, se utilizó ácido salicílico como patrón interno, para evaluar el ajuste de las muestras reales al modelo PLS. La concentración de los analitos en las muestras comerciales fue evaluada empleando el modelo obtenido a través de análisis multivariado de datos en el rango espectral UV. El procedimiento propuesto fue aplicado con éxito en la determinación de sacarina y aspartamo en edulcorantes comerciales.

ABSTRACT

A new, quickly, and low price method to determine simultaneously a mixture of commercial non-caloric sweeteners, as saccharin and aspartame is proposed. A classical full factorial design for standards was used in calibration step to build the partial least square (PLS-2) model. Instrumental data was obtained by mean of UV-visible spectrophotometry. Salicylic acid was used as internal standard in order to evaluate the adjustment of the real samples to the PLS model. The concentration of analytes in the commercial samples was evaluated using the obtained model by UV spectral data. The proposed procedure was applied successfully to the determination of saccharin and aspartame in non-caloric commercial sweeteners.

INDICE

	Página
CAPITULO 1. OBJETIVOS DEL TRABAJO PROPUESTO.....	1
CAPITULO 2. INTRODUCCIÓN	
2.1. Aditivos Alimentarios.....	3
2.1.1. Definición de Aditivos Alimentarios.....	4
2.1.2. Clasificación de Aditivos Alimentarios.....	5
2.1.3. Problemas que causan los Aditivos Alimentarios	
2.1.3.1. Problema Bromatológico.....	5
2.1.3.2. Problema Higiénico-Toxicológico.....	5
2.1.3.3. Problema Analítico.....	6
2.1.3.4. Problema Legal.....	6
2.1.4. Generalidades de los Edulcorantes.....	6
2.1.5. Análisis Instrumental en Edulcorantes Comerciales.....	7

2.2. Metodología de Calibración Multivariada.....	8
2.2.1. Introducción.....	8
2.2.2. Clasificación de los Métodos Multivariados.....	14
2.2.2.1. Clasificación Lineales – No lineales.....	14
2.2.2.2. Clasificación Selección – Espectro completo.....	15
2.2.2.3. Clasificación Directos – Indirectos.....	15
2.2.2.4. Clasificación Inversos – Clásicos.....	16
2.2.2.5. Clasificación Rígidos – Flexibles.....	17
2.2.3 Diferenciación entre Metodologías PLS y PCR.....	17
2.2.4. Fundamentos Matemáticos del PLS.....	20
2.2.5. Tratamiento de Datos Analíticos por PLS.....	26
2.2.5.1. Calibración o confección de un Modelo PLS.....	26
2.2.5.1.1. Raíz Cuadrada del Error Medio de la Calibración y Porcentaje de Varianza Explicada	27
2.2.5.1.2. Raíz Cuadrada del Error Medio de Predicción.....	29
2.2.5.1.3. Otros Parámetros de Evaluación del Modelo.....	30
2.2.5.2. Validación del Modelo.....	31
2.2.5.3. Predicción.....	32
2.2.6. Paquetes Estadísticos para el Empleo de la Metodología PLS.....	33

CAPITULO 3. MATERIALES Y MÉTODOS

3.1. Reactivos.	35
3.1.1. Soluciones Estándar.....	35
3.1.1.1. Solución de Sacarina Sódica.....	35
3.1.1.2. Solución de Aspartamo.....	35
3.1.1.3. Solución de Ácido Salicílico.....	35
3.1.1.4. Solución Reguladora pH2.....	35
3.1.1.5. Agua Deionizada.....	35
3.2. Instrumental	36
3.2.1. Medidas Absorciométricas.....	36
3.2.2. Medidas y Ajuste de pH.....	36

CAPITULO 4. RESULTADOS Y DISCUSIÓN

4.1. Modelo de Calibración obtenido a través de Regresión por Mínimos Cuadrados Parciales (PLS)	37
4.2. Evaluación del Modelo PLS	38
4.3. Evaluación de Sacarina, Aspartamo y Patrón interno en el Modelo de Calibración	41
4.4. Evaluación de efectos de matriz de la muestra sobre el Modelo PLS	43
4.5. Cuantificación de Sacarina y Aspartamo en muestras reales de Edulcorantes	44

CAPITULO 5. CONCLUSIONES

5.1. Del Pretratamiento de Muestra.....46

5.2. De los Costos Instrumentales.....46

5.3. De las Ventajas Multivariadas.....46

5.4. De la Aplicabilidad del Método.....47

CAPITULO 6. BIBLIOGRAFÍA.....48

CAPÍTULO 1. OBJETIVOS DEL TRABAJO PROPUESTO

El uso de herramientas de calibración multivariada, ha ido en creciente aumento, debido a la necesidad de procesar una gran cantidad de datos obtenidos por el instrumental analítico actual y favorecido por el desarrollo de los nuevos sistemas informáticos de gran capacidad de procesamiento (tanto sea programas como equipos y sistemas de computación) generando un conjunto de análisis de datos sumamente valioso para el desarrollo de la Química. Estos sistemas de análisis de datos, han adquirido incluso la categoría de subdisciplina, la cual es conocida en su conjunto con el nombre de Quimiometría. El uso de herramientas de análisis computacionales también se ha aplicado con éxito en otras disciplinas como la Economía, Psicología, Ciencias Sociales, Biología, etc.

La ventaja multivariada radica en la posibilidad de analizar muestras complejas, utilizando herramientas analíticas rápidas y económicas, a través de los sistemas multivariados de análisis. El análisis convencional de muestras complejas implica el uso de procesos separativos engorrosos para la eliminación de interferencias, como así también la utilización de instrumental analítico de elevado costo, el cual no siempre es asequible en laboratorios de docencia.

Un tipo de muestra compleja para analizar resultan los edulcorantes artificiales comerciales. Éstos están normalmente constituidos por combinaciones de diferentes principios activos artificiales: ciclamato, sacarina, acesulfame K y aspartamo. La determinación analítica convencional de estas especies involucra el uso de técnicas instrumentales de muy elevado costo, como por ejemplo la electroforesis capilar, la cromatografía líquida de alta resolución en fase reversa y la cromatografía de intercambio iónico de alta resolución. Si bien estas técnicas instrumentales son referenciales, el elevado costo de equipamiento hace poco factible su adquisición para laboratorios de docencia y de análisis de rutina para control de calidad. Es en este aspecto donde el uso de herramientas multivariadas adquiere connotaciones de importancia y de factibilidad para laboratorios con recursos limitados.

El objetivo del presente trabajo es desarrollar el modelado de un sistema multivariado para la cuantificación de sacarina y aspartamo presentes en edulcorantes artificiales comerciales accesibles en el mercado, empleando absorciometría molecular UV como metodología instrumental y tratamiento de datos espectrales a través de calibración multivariada en su variante PLSR-2 (Regresión por Mínimos Cuadrados Parciales).

CAPÍTULO 2. INTRODUCCIÓN

2.1. ADITIVOS ALIMENTARIOS

El agregado de sustancias de diversa naturaleza en los alimentos, se limitó durante muchos años a los productos naturales, tales como azúcares, gomas, gelatinas, sal, vinagre, colorantes naturales y especias, con el fin de mejorar los alimentos, tornarlos más apetitosos, mejorar su aroma o sabor, la textura, el aspecto general, el color o la estabilidad. Por ser aditivos naturales, eran inocuos o no sospechosos de toxicidad, debido a su uso a través del tiempo.

El progreso de la química, con la posibilidad de sintetizar una multitud de sustancias naturales, o crear infinitas nuevas, impulsó en el siglo XIX el uso de nuevos aditivos. Éstos no siempre fueron utilizados con fines lícitos, sino para aprovechar materias primas deficientes, enmascarar alteraciones o suprimir etapas en los procesos de elaboración y conservación y falsificar diversos productos, presentándolos como otros de mayor calidad, acreditados por una marca. Es por ello que fue necesario realizar controles por parte de los gobiernos, quienes se encontraron con la lógica oposición de industriales y fabricantes, sumado al hecho de existir una amplia diversidad de criterios por parte de bromatólogos y médicos higienistas, respecto a su determinación y efecto en la salud humana respectivamente.

Por otra parte la escasez de alimentos provocada por la disminución de la producción, por guerras, o por aumento de la población, explica como el número de Aditivos Alimentarios ha ido aumentando. Actualmente la JECFA [1], que es el Comité Mixto FAO/OMS de Expertos en Aditivos Alimentarios, ha analizado más de 1500 aditivos.

Los aditivos forman parte del alimento y con él se ingieren, de manera que debe tenerse la certeza de ser inocuos. Además no deben ser empleados para aprovechar materias primas deficientes, ya sea en alteración como en falsificación o adulteración de productos.

2.1.1. Definición de Aditivos Alimentarios

En nuestro país, el Código Alimentario Argentino (CAA) [2], en el artículo 6 define:

“Aditivo Alimentario” como “cualquier sustancia o mezcla de sustancias que directa o indirectamente modifique las características físicas, químicas o biológicas de un alimento a los efectos de su mejoramiento, preservación o estabilización, siempre que:

- a. Sean inocuos por sí mismo o a través de su acción como aditivos en las condiciones de uso.*
- b. Su empleo se justifique por razones tecnológicas, sanitarias, nutricionales o psicosensores necesarias.*
- c. Respondan a las exigencias de designación y de pureza que establece el CAA”*

Las Normas MERCOSUR definen el Aditivo Alimentario de la siguiente manera:

“cualquier sustancia que no se consume normalmente como alimento por sí mismo, ni se usa normalmente como ingrediente típico del alimento, tenga o no valor nutritivo, cuya adición intencional al alimento sea para un fin tecnológico (inclusive sensorial) en la fabricación, elaboración, tratamiento, envasado, empaquetado, transporte o almacenamiento, y que provoque o pueda esperarse razonablemente que provoque (directa o indirectamente), que ella misma o sus subproductos lleguen a ser un complemento del alimento o afecten a sus características”.[2]

2.1.2. Clasificación de Aditivos Alimentarios

Hay numerosos aditivos permitidos en nuestra legislación. En general se los clasifica no en función de su naturaleza ni de sus propiedades particulares, sino de sus propiedades funcionales. Muchos aditivos entran en más de una clase debido a que posee propiedades diversas, así por ejemplo la sacarina es a la vez edulcorante y exaltador del sabor, pudiendo añadirse a un determinado alimento para lograr una o ambas propiedades [2-4].

Los principales grupos de aditivos son:

- Antioxidantes
- Colorantes
- Conservantes
- Edulcorantes no nutritivos
- Emulsionantes
- Gelificantes, espesantes y estabilizantes
- Exaltadores del sabor
- Agentes complejantes de metales

2.1.3. Problemas que causan los Aditivos Alimentarios

2.1.3.1. Problema Bromatológico

El agregado de aditivos, puede llevar a la adulteración de productos, en general, por la sustitución de algunos componentes por otros de menor precio o de inferior valor alimenticio.

2.1.3.2. Problema Higiénico- Toxicológico

La alarma contra el uso indiscriminado de aditivos se originó en las últimas décadas, cuando se descubrió que muchos colorantes sintéticos en uso contenían impurezas muy tóxicas, o bien eran tóxicos por si mismos, especialmente cancerígenos y que varios conservadores químicos provocaban trastornos digestivos graves.

2.1.3.3. Problema Analítico

La gran cantidad de aditivos que pueden presentarse en los alimentos, generalmente en muy pequeñas cantidades, crea un problema analítico para su identificación y evaluación, lo cual ha hecho necesario utilizar todos los recursos analíticos disponibles para resolverlo y establecer límites para los que resulten peligrosos.

En algunos casos, basta con determinar la presencia de un determinado aditivo, para rechazar un producto. Pero en otros, es necesario lograr su valoración en el alimento para saber si excede el límite permitido.

2.1.3.4. Problema Legal

El criterio sobre el uso de aditivos no ha sido hasta ahora uniforme en los diferentes países. A pesar de ello, en las últimas décadas, se han hecho frecuentes las reuniones de carácter internacional, destinadas a aunar criterios, y se ha llegado a coincidir en algunos puntos básicos para la legislación:

- Los alimentos no deben contener sustancias extrañas, a menos que las reglamentaciones lo autoricen expresamente.
- Para que la adición de sustancias extrañas pueda ser autorizada, deben respetarse las siguientes condiciones o requisitos:
 - Ser inocuo para la salud humana;
 - Su empleo debe ser reconocido como necesario.
 - Se debe agregar la cantidad más pequeña posible.
 - No engañar al consumidor, es decir, debe informarse en la etiqueta del producto.
 - Por último, se debe contar con un método analítico adecuado para su identificación y valoración.

2.1.4. Generalidades de los Edulcorantes

Los edulcorantes artificiales son compuestos que tienen la propiedad de otorgar sabor dulce a los alimentos. Son sustancias total o parcialmente sintéticas, con un gran

poder endulzante de entre 10 a 2000 veces superior al de la sacarosa, por lo que su consumo está limitado a cantidades muy pequeñas.

Los edulcorantes artificiales son, por definición, aditivos alimentarios, en oposición a alimentos y están clasificados como tales en el Código Alimentario Argentino [2]. Pueden ser usados separadamente o en combinación, siendo reportados efectos sinérgicos en varias combinaciones [5]. Los edulcorantes más utilizados son: sacarina, ciclamato, aspartamo y acesulfame K, permitiéndose su uso en más de 90 países.

La sacarina, (1,2-benzisotiazolin-3-ona-1,1-dióxido), es un edulcorante artificial de gran poder endulzante (aproximadamente 300 a 500 veces más dulce que la sacarosa). Su uso es aceptado y a las dosis habituales, no se ha demostrado que sea perjudicial para la salud, por lo que se admite su consumo en cantidades moderadas, siendo su ingesta diaria admitida (IDA) de 5 mg kg^{-1} de peso corporal [2, 5-6].

El aspartamo, (N-L-a-aspartil-L-fenilalanina metil éster), tiene un poder endulzante muy alto (200 veces superior al de la sacarosa) y probablemente sea el mejor sustituto de la sacarosa. El aspartamo tiene el inconveniente de que cuando es sometido a altas temperaturas (superiores a $120 - 140^{\circ}\text{C}$), durante un tiempo prolongado, se descompone en ácido aspártico y fenilalanina, perdiendo su capacidad endulzante y produciendo un sabor amargo. La ingesta diaria admitida (IDA) es de 40 mg kg^{-1} de peso corporal [2, 5, 7]. El consumo de aspartamo debe limitarse en personas que padecen fenilcetonuria.

2.1.5. Análisis Instrumental de edulcorantes comerciales

Una gran variedad de métodos analíticos instrumentales han sido aplicados para el análisis de sacarina y aspartamo en alimentos, bebidas livianas, productos dietéticos y preparaciones farmacéuticas. Sin embargo, no existen procedimientos para la determinación simultánea de sacarina sódica y aspartamo mediante el uso de la combinación PLS y Espectrofotometría UV-Visible.

Hoy en día, la Cromatografía Líquida de Alta Resolución (HPLC) es la técnica más frecuentemente usada [8-9]. Además, la Cromatografía de Iones (IC) ofrece una alternativa atractiva a los tradicionales métodos HPLC [10-11].

Hace pocos años atrás, una nueva cromatografía, la Cromatografía Micelar Electrocinética (MEKC) y Electroforesis de Zona Capilar (CZE) han sido aplicadas a la determinación simultánea de varios tipos de Edulcorantes [12-13].

Otros métodos desarrollados para la determinación de sacarina y aspartamo son:

- Espectrometría FT-IR. [14]
- Espectrometría FT-IR con análisis quimiométrico. [15]
- Espectrometría FT-Raman con análisis quimiométrico. [16]
- Amperometría basado en el uso de membranas de bicapa lipídica. [17]
- Espectrometría basada en la complejación de aspartamo con cobre. [18]
- Biosensores para la determinación de aspartamo. [19]
- Electrodo Selectivo de Iones para la determinación de sacarina. [20-21]
- Análisis de Inyección de Flujo (FIA) secuencial acoplada con Detección Enzimática para determinación de aspartamo. [22]
- Análisis de Inyección de Flujo (FIA) con extracción en fase sólida para la determinación de sacarina y aspartamo. [23]

Sin embargo, todos estos métodos están basados en el uso de instrumentos analíticos de alto costo. Por esta razón, en este trabajo, es propuesto un método nuevo, rápido y de bajo costo para la determinación simultánea de dos Edulcorantes Artificiales, Aspartamo y Sacarina Sódica, mediante el uso de Calibración Multivariada de Mínimos Cuadrados Parciales (PLS-2) y Absorciometría Molecular UV.

2.2. METODOLOGÍA DE CALIBRACIÓN MULTIVARIADA

2.2.1. Introducción

El empleo de herramientas multivariadas ha tenido un gran aporte en el estudio de la Química Analítica. Los precursores en estas metodologías analíticas fueron Bruce Kowalsky (Estados Unidos) y Svanden Wold (Suecia) en la década del '80, quienes sentaron precedentes en el tema. A partir de entonces, el desarrollo de nuevas metodologías multivariadas ha tenido un crecimiento relevante, sobre todo a partir de 1990 con el desarrollo de nuevas tecnologías en programas y sistemas de computación. Hoy en día en todo el mundo surgen nuevos trabajos referidos a calibración multivariada para diferentes tipos de muestras y analitos de interés.

En la actualidad, la calibración multivariada es utilizada como herramienta no solo en Química Analítica, sino también su empleo se extiende a otras disciplinas como la economía (econometría), biología (biometría) y diseño de fármacos (QSAR de su sigla en inglés *Quantitative Structure Activity Relation*) [24].

La calibración multivariada es una serie de herramientas matemáticas modernas aplicadas al análisis químico en forma selectiva cuando se utiliza más de una medida de respuesta y más de un analito para cada muestra. Su uso se extiende tanto sea en la cuantificación de constituyentes mayoritarios como en el análisis de microcomponentes, para un gran número de metodologías instrumentales tales como NIR-FT, cromatografía de gases, cromatografía de líquidos y más ampliamente usada en la absorciometría molecular UV-Visible, entre otras [25].

La calibración multivariada es usada para el desarrollo de un modelo matemático-cuantitativo con el objeto de poder obtener una predicción de las propiedades de interés en una muestra (y_1, y_2, \dots, y_q , siendo y , por ejemplo, la concentración de diversos analitos de interés) desde un número de variables respuesta iniciales (x_1, x_2, \dots, x_p , siendo x por ejemplo, medidas de absorbancia a diferentes longitudes de onda). Tanto sea el número de propiedades de interés como el número de variables iniciales óptimas, debe ser determinadas en forma meticulosa, ya que cuando el modelo obtenido es simple (con un bajo ajuste) son introducidos errores sistemáticos, o bien cuando es muy complejo (sobreajustado) se pueden obtener errores aleatorios significativos [26].

El ejemplo más sencillo para el análisis espectrofotométrico de una mezcla con el objeto de medir la concentración de dos o más de sus constituyentes en calibración multivariada, es la regresión lineal múltiple (MLR de su sigla en inglés). En esta metodología se eligen un número de longitudes de onda (variables x) proporcionales al número de analitos de interés (variables y) y en donde la absorbancia en las longitudes de onda seleccionado sea máxima para un componente y mínima para el resto. En este caso, la resolución de la mezcla puede realizarse manualmente resolviendo un sistema de ecuaciones con n variables respuesta y $n-1$ variables de interés.

En términos generales, la calibración uni o multivariada reemplaza la medida de la propiedad de interés por una que es más accesible y suficientemente exacta a través del uso de modelos analíticos y matemáticos adecuados. Un modelo de calibración requiere la formulación de los objetivos de estudio, adecuado diseño experimental, selección del modelo con mejor ajuste, estimación de sus parámetros y de la precisión de las predicciones.

En calibración uni o multivariada el procedimiento analítico general esta constituido por dos etapas:

- Tratamiento de la muestra: en donde se deben realizar algunos procedimientos químicos y/o físicos de diversa complejidad a fin de obtener el acondicionamiento de la muestra y del sistema, donde una cantidad física medible es correlacionada con una cantidad química a ser medida.
- Cuantificación: El proceso de cuantificación en calibración uni o multivariada consta de tres etapas [25]:
 - a. *Modelado o calibración*: en la que se estudia el comportamiento del método con el objeto de encontrar una función matemática que describa el modelo de dicho comportamiento ($y = f(x)$), a partir de dos conjuntos de variables donde y es la variable dependiente y x , la independiente. En la etapa de calibración, la serie de datos utilizados aquí se conoce como conjunto o “set” de calibración y los parámetros del modelo se conocen como coeficientes de regresión o sensibilidades. El uso de un modelo de calibración puede basarse en la aplicación de una relación teórica establecida (la ley de Lambert-Beer en espectrofotometría, la ecuación de Ilkovic en polarografía, la ecuación de Nernst en potenciometría, etc.) o bien puede resultar de una aproximación empírica, surgida de la experiencia del analista. En los métodos instrumentales de análisis la calibración es una función obtenida a través del ajuste entre un modelo matemático con un conjunto de datos obtenidos experimentalmente. Por inversión de la función de calibración se obtienen las variables de interés.
 - b. *Predicción*: consiste en obtener las variables dependientes de las muestras de interés que permiten predecir el valor de las variables independientes conocidos los coeficientes de regresión. En la etapa de predicción el conjunto de datos usados en ella constituye el conjunto o set de predicción.
 - c. *Validación*: en donde la operación de predicción se realiza con un conjunto de muestras de composición conocida y el nuevo grupo se conoce como conjunto o set de validación. Esta etapa es de utilidad a fin de conocer el error asociado al modelo en la etapa de predicción y evaluar la capacidad del mismo.

En los casos en que el instrumento registra la señal en función de una sola variable (instrumentación con metodologías clásicas) la calibración resulta sencilla y consiste en la búsqueda de una función matemática que vincule el valor de la medida (respuesta instrumental) con el valor de la variable (concentración) para un conjunto de muestras de composición conocida (sistema univariado). Una recta constituye la función de calibración más simple y frecuentemente utilizada, donde los parámetros de la ecuación de la recta son las sensibilidades que definen los parámetros del modelo.

Sin embargo, la moderna instrumentación analítica en la actualidad permite obtener en un período de tiempo breve una gran cantidad de información, como por ejemplo la fluorescencia molecular, absorciometría molecular y diferentes técnicas acopladas como la cromatografía líquida con detector de diodos en línea, HPLC-DAD y la cromatografía de gases con detector de masas, GC-MS, que permiten obtener conjuntos de datos en función de dos o más variables (sistemas multivariados) [25].

El uso frecuente de una sola señal o un solo valor de la variable independiente produce una importante pérdida de información relevante. Pero el uso de un número importante de variables independientes permite obtener mucha más información sobre la muestra si el conjunto de datos es tratado adecuadamente mediante el uso de herramientas multivariadas.

Los métodos de calibración multivariada involucran a más de una variable independiente ($p > 1$), a diferencia de la calibración univariada en donde solamente hay una única variable ($p = 1$). Esta diferencia abre nuevas posibilidades como por ejemplo el uso de todas o gran número de longitudes de onda en un espectro completo como mecanismo de calibración y predicción, en vez de la señal simple a la longitud de onda más sensible. El uso de información espectral completa puede permitir mejores predicciones, pero también existe la posibilidad de que regiones espectrales poco o nada relevantes sean incluidas en el modelo durante la calibración. Sin embargo el uso de más de una variable independiente permite la estimación de más de una propiedad dependiente ($q > 1$) y la corrección de covarianzas indeseables (interferencias). En definitiva, esta es la mayor motivación para realizar mediciones multivariadas.

En calibración univariada no es posible corregir interferencias sin información adicional, pero con datos multivariados, la oportunidad surge al separar la información relevante para las respectivas propiedades, de las variaciones no relevantes o ruidos aleatorios [26].

En calibración multivariada, la precisión del resultado final no depende tanto de la precisión de los datos o variables originales, ya que a veces la medida absoluta o precisa de cada variable individual puede ser irrelevante para el propósito del análisis. En la espectroscopía de reflectancia difusa en infrarrojo cercano (NIRS) no importa obtener lecturas de absorbancias aceptadas universalmente a las mismas longitudes de onda individuales, sino que el fin último es predecir el contenido de proteína en cierto tipo de muestras de trigos. Por ello, si bien la reflectancia o transmisión NIR desde las muestras intactas parecen sumamente confusas, la calibración multivariada convierte varias longitudes de onda útiles en información precisa y relevante.

En ciertas aplicaciones, la calibración multivariada no surge como un detallado entendimiento de la relación entre variables dependientes e independientes, por lo que la habilidad para interpretar un modelo no es un requerimiento ni condición necesaria. Esto no impediría la exploración de conocimientos previos accesibles sobre el problema durante el modelado de la calibración. Un modelo que puede ser adecuadamente interpretado ciertamente agrega valor y confianza al resultado de la calibración [26].

El número de posibles aplicaciones de modelado predictivo es virtualmente ilimitado. La mayor aplicación actual se centra en el área de la química analítica, sobre todo en el desarrollo y aplicación de modelos de calibración predictivos con fines cuantitativos. Tal es el caso de la determinación simultánea de la concentración de varios analitos en una mezcla multicomponentes donde se puede elegir la metodología más adecuada a partir de un gran arsenal de métodos espectroscópicos (por ejemplo, UV-Visible, IR, NIR, XRF, NMR). Por otra parte, el uso de sensores múltiples en línea resulta ser otra posibilidad muy interesante y depende casi exclusivamente de la aplicabilidad de modelos de calibración multivariada para el monitoreo cuantitativo de los sistemas o procesos de interés. Por ejemplo, la aplicación de espectroscopía de infrarrojo cercano para analizar pequeñas muestras en sistemas en línea o sistemas discontinuos ha encontrado un uso extendido en las industrias químicas y en las industrias de alimentos. La metodología es útil para caracterizar e identificar propiedades de productos relacionados a su composición, cuyo análisis por metodologías clásicas puede resultar costoso, por ejemplo determinación del número de octanos en combustibles, valores de yodo y grado de insaturación en grasas y aceites o cristalinidad de polímeros.

Otras aplicaciones por fuera del campo de la química analítica son la predicción de propiedades farmacológicas o bioquímicas desde parámetros estructurales (QSAR), la comprensión de perfiles sensoriales y organolépticos desde datos fisicoquímicos en investigaciones de alimentos y la obtención y uso de modelos a partir de datos ambientales [26].

La calibración multivariada puede estimular el desarrollo de nuevo instrumental analítico, a través de su incorporación a nuevos y más sofisticados equipos, como así también incrementar la capacidad analítica y la precisión de los instrumentos tradicionales actuales.

La aplicación de metodologías de calibración multivariada al tratamiento de datos analíticos requiere que el analista conozca de los fundamentos matemáticos, como así también del sistema analítico de estudio, para que en conjunto los resultados sean comparables con otras metodologías.

Los sistemas de calibración multivariada mas utilizados en la Química Analítica son:

- Análisis de Componentes Principales (PCA de *Principal Components Analysis*)
- Regresión de Componentes Principales (PCR de *Principal Components Regression*)
- Regresión por Cuadrados Mínimos Parciales (PLS de *Partial Least Square Regression*)

De los anteriores, el más frecuentemente utilizado en Química Analítica para fines cuantitativos es el PLS. El PCA se lo utiliza a menudo para clasificación de conjuntos de muestras, a partir de sistemas predefinidos, por lo que su uso se limita al análisis cualitativo. Tal es el caso de clasificar vinos utilizando como variables caracteres organolépticos y fisicoquímicos.

El PCR se puede utilizar para cuantificación de analitos, pero en sistemas reales puede tener el inconveniente de no considerar adecuadamente interferencias desconocidas, ya que considera los errores aleatorios de todo el sistema en conjunto (de la matriz de respuesta y de la matriz de concentraciones). Como ejemplo, el PCR puede utilizarse en sistemas sencillos y conocidos y como herramienta en sistemas en línea para control de calidad en la industria, en donde no ha de esperarse la incorporación de sustancias interferentes o cambios importantes en la composición que puedan afectar la variabilidad del sistema en el modelo.

El PLS tiene la ventaja frente al anterior, de tener en cuenta interferencias que no hayan sido previstas durante la confección del modelo, ya que evalúa en forma independiente los errores de la matriz de respuesta por un lado y los de la matriz de concentraciones por el otro, lo que le da una mayor versatilidad debido a que es más factible su aplicación a sistemas reales. El PLS se ha aplicado a un amplio rango de muestras para determinar diversos analitos de interés en ellas [27-38].

2.2.2. Clasificación de los Métodos Multivariados

Como se explicó anteriormente, las metodologías de calibración multivariada presentan diferencias que pueden ser notables tanto sea a los requisitos del modelo a aplicar y a las exigencias del cálculo implicado en el proceso de calibración [25].

La selección de un dado método dependerá del tipo de problema, del conocimiento previo del sistema y de los datos disponibles.

Es posible realizar la siguiente clasificación entre los métodos de calibración multivariada:

- Lineal - No lineal
- Selección - Espectro completo
- Directa - Indirecta
- Inversa - Clásica (reversa)
- Rígidos - Flexibles

2.2.2.1. Clasificación Lineales – No lineales

La diferencia fundamental entre sistemas lineales y no lineales es la existencia o suposición de funciones lineales o no lineales entre las variables X. Un método de calibración lineal puede escribirse de la siguiente manera:

$$y = b_0 + \sum_{i=1}^n b_i x_i + e$$

Ecuación 2.1.

donde los parámetros b_0 y b_1 pueden ser determinados a partir de los datos del conjunto de calibración.

Un ejemplo de calibración en un método no lineal es la ecuación 2.2.:

$$y = a x^b + e$$

Ecuación 2.2.

Sin embargo, cuando un sistema de este tipo no es el apropiado, se puede realizar un preprocesamiento no lineal y transformarlo en un modelo lineal.

2.2.2.2. Clasificación Selección – Espectro completo

En calibración multivariada una de las decisiones más importantes es la de utilizar todas las variables X respuesta posibles o solo unas pocas. Dependiendo de la complejidad del sistema a analizar, algunos de los métodos usados en la práctica pueden emplear un pequeño número de variables e ignorar el resto, por razones puramente de cálculo. Por otra parte los métodos de espectro completo pueden usar todas las variables disponibles significativas.

La distinción entre selección y espectro completo es importante cuando las variables son altamente colineales. En este caso, la utilización de todas las variables puede ser inútil y resulta más adecuado reducirlas en forma significativa, ya que en este caso hay información reiterativa o repetitiva en muchas de las variables iniciales.

2.2.2.3. Clasificación Directos – Indirectos

La distinción entre ambas es determinada en función de la información previa. En la calibración directa los parámetros del modelo son conocidos a priori y se pueden usar para definir la función en situaciones más bien ideales. Un ejemplo de esta situación puede ser la resolución espectrofotométrica de una mezcla de componentes conocidos a partir de los espectros de los componentes puros. Pero sin embargo la presencia de interferentes, ruido o respuesta no lineal, de interacciones entre los constituyentes, etc., limita cualquier posibilidad de uso. Por ello su utilización queda restringida a sistemas puntuales ideales o casi ideales.

En algunas situaciones no es posible obtener los espectros de todos los componentes en una disolución. En otras, algunos constituyentes pueden modificarse a través de reacciones transformando sus características. Además puede ocurrir que algunas interferencias no pueden aislarse o simplemente se desconoce por completo su existencia. En todos estos casos, la calibración indirecta ofrece una alternativa más viable y factible de aplicar. Este es un tipo de calibración más empírico que requiere un conjunto de patrones de concentración conocida de los constituyentes a determinar. Estos datos se utilizan para generar un modelo que permita realizar la predicción de las características del componente de interés, considerando la presencia de interferencias en el mismo modelo. Para ello no es necesario conocer ni los espectros ni las concentraciones de las distintas interferencias. La única condición es que en la calibración los patrones deben elegirse o prepararse de tal manera a fin de que en ellos estén representados lo mejor posible todos los interferentes.

Dado que en este caso se realiza una estimación estadística de los parámetros desconocidos del modelo de calibración (interferencias) a partir de mezclas de composición conocida, se requiere una mayor cantidad de datos que la anterior (mayor cantidad de patrones y la mayor cantidad de variables independientes relevantes posible).

Puede también existir una situación intermedia en que algunos parámetros son conocidos a priori y otros se determinan a partir de los datos de calibración.

2.2.2.4. Clasificación Inversos – Clásicos

La diferencia entre ambas se establece según el diseño del modelo de calibración. Ello depende de si se define una ecuación de Y en función de X o bien de X en función de Y .

La notación clásica para expresar un modelo para un conjunto de objetos esta representada en la ecuación 2.3.:

$$X = YK + E$$

Ecuación 2.3

donde X representa la matriz de datos espectrales, Y la matriz de concentraciones de los analitos, K es la matriz de espectros unidad de los analitos y E representa los errores del modelo y ruido aleatorio. Si K no es conocido previamente se puede determinar de una manera indirecta.

Este tipo de modelos constituyen los sistemas más básicos de calibración y de ahí su distinción de clásico. Sin embargo, si el principal objetivo de la calibración es realizar la predicción de la matriz Y , parece más lógico y adecuado definir un modelo de calibración inversa en donde se defina Y en función de X . En este modelo se realiza una regresión de los datos de Y sobre las variables X , como se muestra en la ecuación 2.4:

$$Y = X B + E$$

Ecuación 2.4.

donde X representa la matriz de datos espectrales, Y la de concentraciones de los analitos de interés, B es la matriz de los coeficientes de regresión y E la de los residuales (errores asociados al modelo, ruido instrumental, errores operacionales, etc.).

2.2.2.5. Clasificación Rígidos – Flexibles

Pueden surgir por combinación de características de los ya citados. En los primeros, el modelo definido por el número de variables usadas (patrones, muestras, componentes del sistema, nivel de ruido, etc.), queda rígidamente establecido y solo funcionan correctamente cuando se cumplen todas las premisas del mismo. La incorporación de una nueva fuente de variación no considerada durante el modelado puede causar graves distorsiones y errores considerables en la predicción. En los modelos rígidos se realiza una regresión de las variables sobre tantos patrones como analitos hay en la muestra.

En los métodos flexibles, por el contrario, se realiza la regresión de las variables sobre un número de patrones que puede ser distinto al de analitos de interés, por lo que estos métodos son más estables y están menos sujetos a distorsión o error por la presencia de otra fuente de variación, pudiendo soportar pequeñas variaciones sin afectar significativamente la predicción [25-26].

2.2.3 Diferenciación entre Metodologías PLS y PCR

Los cuadrados mínimos parciales (PLS de *Partial Least Square*) [39-40] y la regresión de componentes principales (PCR de *Principal Component Regression*) son los métodos de calibración multivariada más ampliamente usados en Química Analítica.

Ambos hacen uso de la calibración inversa [41] en donde es posible calibrar los componentes deseados mientras implícitamente se modelan las otras fuentes de variación. La estimación del modelo de calibración inversa involucra la inversión de una matriz típicamente inestable. Tanto el PLS como el PCR resuelven el problema de esta inversión reemplazando las variables originales con combinaciones lineales de dichas variables (factores).

La diferencia entre PLS y PCR radica en cómo son calculados estos factores [25, 39]. Dadas las dos aproximaciones para resolver el problema de la calibración inversa, las ventajas sobre la regresión lineal múltiple (MLR de *Multiple Linear Regression*) son varias. El MLR puede generar modelos sencillos pero limitados por el poco número de variables que es capaz de utilizar. Las ventajas de usar más variables son:

- El modelo resulta ser más sensible a la detección de muestras que son inusuales comparadas con el conjunto de calibración (puntos discrepantes o outliers) [25], lo cual permite descartar a priori tales puntos anómalos.
- Las técnicas de espectro completo (uso de todas las variables posibles) poseen la ventaja multivariada [25, 39]. Esto está relacionado a la señal promediada, donde la desviación estándar es reducida por un factor de \sqrt{n} cuando el promedio de n determinaciones es usado en lugar de una simple medida. Con datos multivariados la señal promediada toma lugar cuando alguna medida cercana redundante es usada para construir el modelo [39, 42]. Esto es análogo a la cuantificación en cromatografía, ya que la determinación de la concentración de cierta especie a través de la medida del área del pico, se opone a la relacionada a la altura del mismo [40]. Esto se debe a que la precisión del área es usualmente mucho mejor que la precisión de la intensidad en un dado punto.

En la metodología PLS existen dos aproximaciones que pueden ser usadas para desarrollar un modelo de calibración para los analitos de interés. Usando el término PLS-1, se generan modelos separados para cada analito. Con PLS-2, un solo modelo se construye usando todos los analitos simultáneamente. En este último caso se deben elegir adecuadamente las condiciones y número de componentes PLS a utilizar en el modelo final [25- 26, 39].

Tanto sea para PLS-1 como para PLS-2 en los modelos resultantes están implícitos todas las interacciones que se producen dentro del sistema, como la competencia de los analitos para la formación de un dado complejo, efecto de interferentes, ruido o fondo instrumental, etc.

El utilizar PCR o PLS dependerá del analista y de su inclinación por uno u otro método. Muchos estadísticos prefieren PCR porque está estadísticamente bien estudiado y definido, mientras que muchos científicos no estadísticos encuentran la aproximación PLS más fácil de entender conceptualmente. También muchos quimiométricos y otros analistas prácticos de datos prefieren PLS porque es efectivo y provee resultados que son más fáciles de interpretar por ser menos complejos, debido a la utilización de solo unas pocas componentes o factores PLS.

A menudo PCR puede dar errores de predicción tan bajos como el PLS, pero usando más componentes para confeccionar un modelo. Sin embargo la solución PCR es más estable, ya que el error en la predicción se aplana alrededor del óptimo y no comienza a incrementarse debido al temprano sobreajuste. PLS es más sensible al sobreajuste, por lo que se torna crucial encontrar el número óptimo de componentes PLS.

PLS-2 es un método natural para comenzar cuando hay muchas variables de respuesta X (de aquí en adelante la definiremos como variables R). Rápidamente da una visión general de los patrones base y se observa si hay correlación significativa entre las variables R . También PLS-2 puede conducir a mejores resultados si R es colineal, porque utiliza toda la información disponible en R . Sin embargo, esta no es una situación muy común. El inconveniente es que se pueden necesitar diferentes números de componentes para las diferentes variables R , lo que debe recordarse para la interpretación y predicción. Este es también el caso para PCR con algunas variables R , dado que cada una puede necesitar muchos o pocos componentes.

Hay casos donde PLS-2 fracasa al intentar modelar algunas de las variables R . En esos casos se debe probar modelos PLS-1. Definitivamente se necesita interpretar cada modelo separadamente [26].

Finalmente se puede concluir que PLS probablemente dará resultados más rápidos, pero lo aconsejable es probar ambos PLS-1 y PLS-2. Este último es utilizado para una primera investigación, pero a menudo se necesitará del modelo PLS-1 separado para conseguir errores de predicción más satisfactorios [25-26, 43].

2.2.4. Fundamentos Matemáticos del PLS

El PLS es un procedimiento de modelado que estima simultáneamente los factores subyacentes en la matriz de concentraciones o variables dependientes Y (de aquí en adelante la definiremos como variables C) y en la matriz de variables independientes o respuestas R . Estos factores son usados para definir un nuevo subespacio en R que mejor ajuste a C .

En PCR, la rotación definida por los autovalores (scores) se utiliza para encontrar un subespacio en R que subsecuentemente modela a C . El PLS es similar a PCR, con la diferencia que los factores son elegidos para tanto las variables en C como en R . Esto se realiza utilizando las columnas de la matriz C para estimar los factores en R . Simultáneamente, las columnas de R son utilizadas para estimar los factores en C .

El PLS responde a un método inverso [39, 44] cuya característica principal es por lo general, que el modelo permite relacionar las medidas instrumentales y las concentraciones. La concentración se calcula en función de las respuestas, como lo muestra la ecuación 2.5 [25, 40-41].

$$R = TP + E \text{ (a)}$$

$$C = UQ + F \text{ (b)}$$

Ecuación 2.5.

donde los elementos T y U son matrices llamadas *scores* de R y C respectivamente y relacionan la información en la matriz de respuesta R con la matriz de concentraciones C , a través de la correlación que existe en las matrices de covarianza de R y C (o sea máxima elongación de los datos originales sobre el nuevo subespacio originado por los componentes PLS). Los elementos P y Q son matrices llamadas *loadings* de R y C respectivamente, que indican las coordenadas de los datos iniciales sobre las nuevas componentes PLS. Un gráfico de los loadings en función de las longitudes de onda, permite observar cuales de ellas son relevantes en el modelo y, adicionalmente, se pueden tener curvas de loadings similares a los espectros, sobre todo cuando no existen fuentes importantes de variación. E y F son las matrices de error asociado al modelado de R y C para el modelo PLS.

Los scores T no son óptimos para estimar las columnas en R , pero son rotados a fin de poder describir simultáneamente la matriz C . El cálculo a través de herramientas PLS también da como resultado una matriz auxiliar conocida como matriz de pesos (W o *loadings weight*), que expresa la correlación entre las matrices U y R y es usada para calcular la matriz T .

En situaciones de idealidad, las fuentes de variación de R son exactamente iguales a las fuentes de variación en C , siendo por consiguiente idénticos los factores T y U para R y C . En aplicaciones reales, R varía en forma no correlacionada a la variación de C (ya que R depende de variaciones instrumentales y C de variaciones operacionales), por lo que $t \neq u$ (escritos ahora en letra minúscula porque representan los vectores que dan origen a las coordenadas de cada uno de los componentes PLS utilizados en el modelo). Sin embargo, cuando las matrices R y C son utilizadas para estimar los factores, existen entre ellos una nueva relación, definida por la ecuación 2.6.:

$$u = bt + g$$

Ecuación 2.6.

donde b es el término de relación interna entre u y t y g el error asociado. b es usado para calcular los factores subsecuentes en R . Si R tiene j vectores columna y C tiene k vectores columna, el vector t estará cerca de los j vectores de R y será un buen estimador para R . Por la misma razón, el vector u estará cerca de los k vectores de C y será un buen estimador para C . Si existe un buen modelado, que vincule a R y C , una gráfica de u vs. t para cada componente principal, deberá dar una línea recta que pase por el origen y por consiguiente, un valor de g muy pequeño. La figura 2.1. muestra como funciona el PLS teniendo en cuenta las anteriores consideraciones.

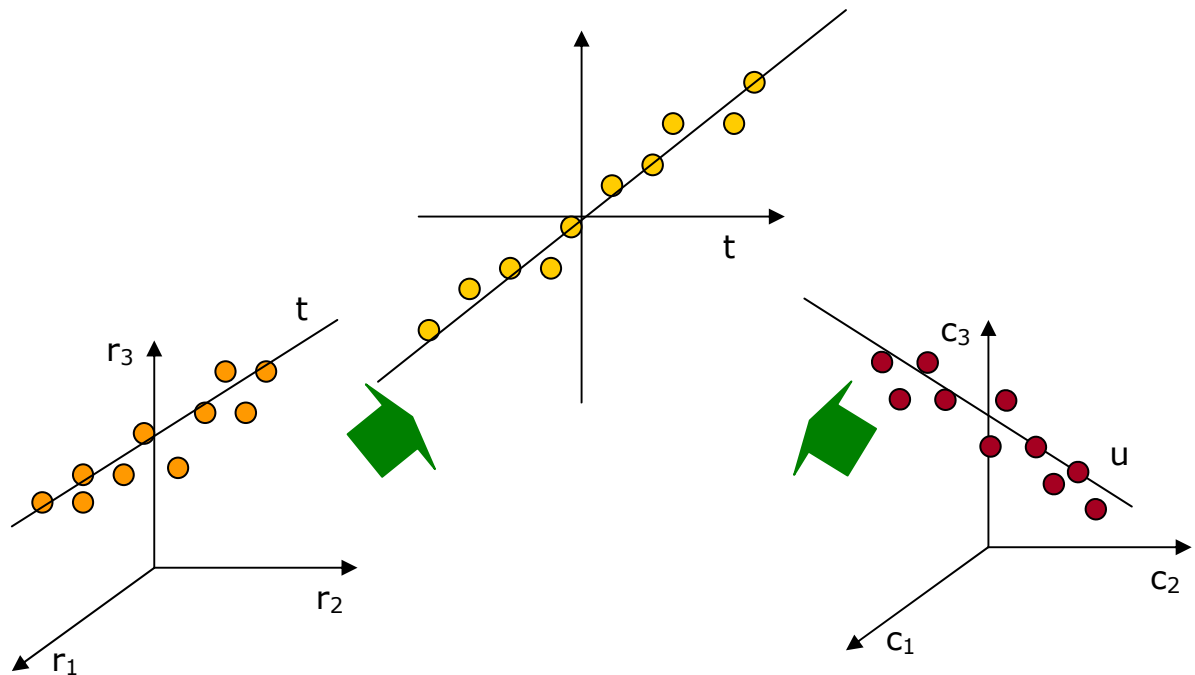


Figura 2.1. Descripción geométrica de operación para metodología PLS.

En la figura 2.1. se aprecia el funcionamiento de un sistema PLS para un caso simple con una matriz de respuesta con tres variables (p.e. absorbancias a 3 longitudes de onda) y una matriz de concentración con tres analitos de interés. La información de ambas matrices es luego condensada por los vectores t y u , las cuales por lo general presentarán una función lineal entre sí [41, 43]. Una vez obtenido el modelo matemático respectivo, a través del PLS se puede predecir los parámetros de interés (concentraciones de analitos) en una muestra incógnita. Dicha muestra producirá un vector de respuestas r_{desc} , el cual es operado matemáticamente en la ecuación 2.5 (a) para obtener un vector t_{desc} , con el cual se puede conocer con la ecuación 2.6 el vector u_{desc} , que finalmente es utilizado para conocer la matriz de concentraciones de la muestra incógnita c_{desc} utilizando la ecuación 2.5 (b). En definitiva, el modelo estará conformado por una serie de matrices y vectores ordenados por el programa, que operan matricialmente con los datos de respuestas o señales de la muestra incógnita.

El PLS opera matemáticamente a fin de que las direcciones de los primeros componentes PLS (definidos por t y u) puedan explicar la mayoría de la variación de la muestra, siendo esta variación la información relevante de la misma necesaria para el modelado. Para ello, PLS utiliza las covarianzas del sistema y son las que definirán la dirección de máxima elongación de los puntos y la dirección de los componentes PLS.

La consecuencia de esto es que la información inicial contenida a lo largo de n variables (por ejemplo valores de absorbancias de un barrido espectral completo), se ve condensada en unas pocas nuevas variables ortogonales entre sí, con lo cual queda resaltada solamente la información útil con los primeros componentes PLS y la información redundante queda confinada en los últimos (definida por la matriz de errores). Si se seleccionan un número elevado de componentes PLS, el modelo se ajustará adecuadamente pero con el costo de tener incorporado en él gran cantidad de información irrelevante (sobreajuste), con lo cual las matrices de error se hacen pequeñas al utilizarse la información relevante e irrelevante. Por el contrario, si el número de componentes PLS seleccionado es muy pequeño, se corre el riesgo de dejar de lado información vital del sistema en estudio (subajuste) y las matrices de error del modelo se tornan grandes [26, 43].

En la figura 2.2. se representa el funcionamiento del PLS en términos generales para hallar los componentes PLS.

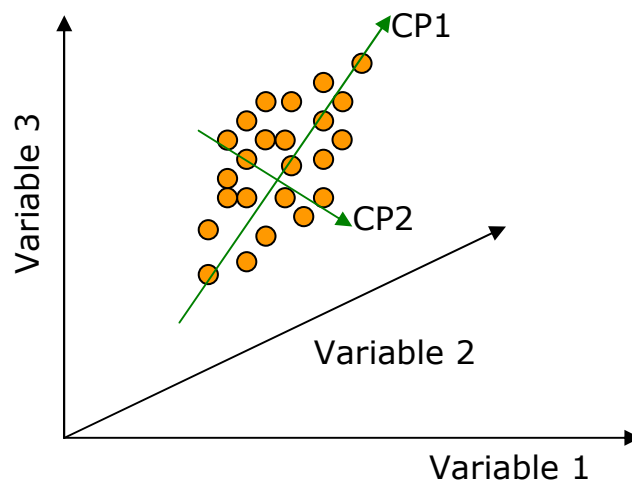


Figura 2.2. Dirección de los componentes PLS a partir de una nube de puntos tridimensional.

En el gráfico se observa claramente como opera el PLS para obtener un modelo a partir de una nube de puntos tridimensional. Se puede observar que el primer componente principal (CP1) tiene la dirección hacia la máxima elongación de la nube de puntos en las tres dimensiones del ejemplo (variables respuesta originales), definida por la matriz de covarianzas del sistema. El segundo componente principal tiene la dirección de la segunda mayor elongación de la nube en las tres dimensiones, ortogonal al CP1.

Cada componente principal que subsecuentemente surja en n dimensiones, será ortogonal a los anteriores, por lo que se generará un nuevo sistema de ejes cartesianos que se orientará hacia donde exista la mayor elongación de la nube de puntos y por ende, mayor información. Por otra parte, se produce una sustancial reducción de variables ya que, como se aprecia en el gráfico, dos componentes PLS modelan a las tres variables originales y esto se hace más sensible en espacios n dimensionales con n variables con una reducción a unos pocos componentes PLS.

A continuación se expondrán algunos conceptos matemáticos de interés para un mejor entendimiento de la metodología PLS.

- a. *Varianza*: de una variable resulta una medida de la dispersión de los valores variables respecto a la media o cuan grande es el rango de valores que pueden tomar (ecuación 2.7.). Este concepto es de crítica importancia para pensar como llegar a los resultados.

$$\text{Var}(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Ecuación 2.7.

La raíz cuadrada de la varianza, la desviación estándar, que es generalmente usada, es también útil para expresar la medida de esta expansión en las mismas unidades de las medidas (ecuación 2.8.) [43]

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Ecuación 2.8.

- b. *Covarianza*. La covarianza entre dos variables, x_1 y x_2 es una medida de su asociación lineal. Si muchos valores de la variable x_1 se vinculan con muchos valores de la variable x_2 la covarianza será positiva. Y viceversa, si muchos valores de la variable x_1 se vinculan con pocos de la variable x_2 , la covarianza será negativa. Una gran covarianza (en valores absolutos) significa que existe una fuerte dependencia lineal entre las dos variables, mientras que si la covarianza es pequeña, luego las dos variables no son muy dependientes; esto es, si cambia la variable x_1 , esto no afecta mucho a la variable x_2 (ecuación 2.9.).

$$\text{COV}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Ecuación 2.9.

La medida de la covarianza depende de las unidades de las variables. Por lo tanto, para comparar por igual dependencias lineales, se utiliza otra medida más práctica, la correlación [43].

- c. *Correlación.* La correlación entre dos variables es calculada por división de la covarianza con el producto de sus respectivas desviaciones estándares, siendo simplemente una medida sin unidad de covarianza escalada. Es en general la más usada para comparar la independencia entre variables, debido a que las correlaciones son directamente comparables en cualquier unidad en que las variables son medidas. El coeficiente de correlación de Pearson, r , está definido por la ecuación 2.10.

$$r = \frac{\text{COV}(x, y)}{S_x S_y}$$

Ecuación 2.10.

donde $\text{cov}(x,y)$ es la covarianza entre las variables x e y . Siendo S_x y S_y las correspondientes desviaciones estándares de las mismas.

El valor r^2 es usado muy a menudo como una medida de la fracción de la varianza total que puede ser modelada por esta medida de asociación lineal.

Los valores de correlación siempre se hallan entre -1 y 0 , y entre 0 y $+1$. Una correlación de 0 , significa que no hay correlación, no hay relación entre las variables; una correlación de $+1$ significa que hay una relación máxima positiva lineal entre las variables, y en una correlación -1 la relación es negativa e inversa [43].

Para hallar la máxima elongación en la nube de puntos n dimensional de las variables originales, el PLS utiliza las covarianzas de las medidas de C y R para obtener los vectores t y u que definirán posteriormente la orientación de los componentes PLS.

El valor r^2 es un índice muy importante que definirá la calidad del ajuste del modelo PLS obtenido.

Como ya se dijo anteriormente, la ventaja de los modelos inversos (tal el caso del PLS), es que es capaz de predecir la concentración de una dada especie cuando existen fuentes de variación química y física presentes en el sistema. Es por ello que esta metodología fue empleada en el presente trabajo como herramienta multivariada para la cuantificación de sacarina y aspartamo en muestras reales de edulcorantes comerciales.

2.2.5. Tratamiento de Datos Analíticos por PLS

El análisis de datos analíticos por metodología PLS requiere necesariamente de tres etapas:

- Confección del modelo o modelado (calibración)
- Validación del modelo
- Predicción (de muestras incógnita)

Los requerimientos del sistema para que la predicción sea posible es que dicho sistema de medida esté adecuadamente diferenciado de otras fuentes de variación, que la respuesta del instrumento sea lo suficientemente lineal con la concentración y que las experiencias de calibración y generación del modelo sean cuidadosamente planeadas [25, 40].

2.2.5.1. Calibración o confección de un modelo PLS

Como ya se comentó anteriormente, un modelo PLS es obtenido a partir de un determinado número de calibradores o patrones, los cuales contendrán en cantidades conocidas y variables, los distintos analitos de interés que se espera hallar en la muestra real. Si bien el PLS es capaz de generar modelos en los cuales los efectos por interferencias puedan estar implícitamente incorporados en él, lo más aconsejable es preparar los patrones de calibración con una composición lo más similar posible a la de la muestra, a fin de evitar resultados dudosos o engañosos.

Para la confección del modelo PLS existen ciertas herramientas de diagnóstico que son usadas para investigar ciertos aspectos de las series de datos: el modelo, los patrones y las variables.

El uso primario de la herramienta de diagnóstico de modelo sirve para determinar el número óptimo de componentes PLS del modelo.

La herramienta de diagnóstico de patrones es usado para estudiar las relaciones entre ellas e identificar posibles puntos discrepantes, debidos a errores de diversa índole (en la matriz de concentración, matriz de respuestas por ruido no modelado, etc.).

La herramienta de diagnóstico de variables es exactamente lo mismo que el anterior, pero inherente a las variables.

Los puntos discrepantes u outliers son patrones o variables atípicas. Si son resultados de medidas erróneas deben ser quitados del modelo, o éste no será el correcto. Pero por otro lado pueden ser valiosos si representan un fenómeno importante desconocido y en ese caso hay que conservarlos; el eliminarlos podría implicar un modelo incorrecto. La decisión final corre por cuenta exclusiva del analista. El problema de identificar puntos discrepantes puede resultar difícil. Los gráficos de % varianza explicada pueden ser usados con este propósito. Esto puede involucrar mucho trabajo, especialmente cuando hay muchas variables y/o patrones. Los gráficos de scores t en función de u son también apropiados para la detección de puntos discrepantes. El patrón en cuestión aparecerá muy separado del resto, con valores excesivamente altos o bajos.

A continuación se detallan las herramientas de evaluación anteriormente nombradas [40].

2.2.5.1.1. Raíz cuadrada del error medio de la calibración y porcentaje de varianza explicada

El primer diagnóstico para la evaluación de un modelo es el porcentaje de varianza explicada para los datos de concentración y medidas. El porcentaje de varianza explicada es la cantidad de varianza que define el modelo con un dado número de factores relativo a la varianza total del sistema. Este diagnóstico sirve para evaluar problemas con los datos, tales como errores significativos en los valores de concentración de referencia o desviaciones en la ley de Beer entre los espectros y las concentraciones.

Cuando algo de esto sucede, el diagnóstico surge por el hecho de que el porcentaje de varianza para la matriz de concentraciones C , es mucho menor que para la matriz de respuesta R para el mismo número de factores. Otro uso útil es ayudar a determinar el número de factores a utilizar en el modelo por comparación entre el porcentaje de varianza remanente y un nivel de ruido conocido en los datos espectrales y/o de concentración definidos por las matrices de error. El porcentaje de varianza explicada se calcula a partir de la raíz cuadrada del error medio de la calibración, RMSEC (de *Root Mean Square Error Calibration*) (ecuación 2.11.) y el factor r^2 que es el factor de regresión del modelo. El RMSEC debería disminuir a medida que se aumenta el número de componentes PLS utilizados para el modelo. Sin embargo, si el número de dichos componentes es demasiado elevado, puede incluir en ellos información irrelevante asociada con los ruidos del sistema.

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_{i,\text{cal}} - y_{\text{cal}})^2}{n}}$$

Ecuación 2.11.

La figura 2.3. muestra un gráfico típico de porcentaje de varianza explicada acumulativa en función del número de componentes PLS, en donde se observa una tendencia máxima a partir del componente número 3 (en el ejemplo hipotético), lo que significa que el modelo está utilizando un alto porcentaje de información de los datos originales a partir del componente 3 [40, 43].

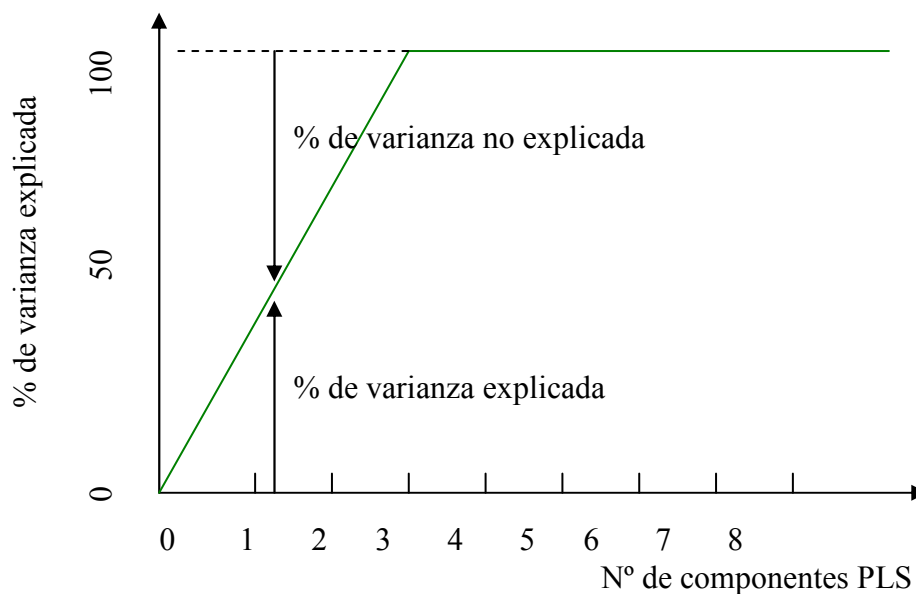


Figura 2.3. Porcentaje de varianza explicada en función del número de componentes PLS.

2.2.5.1.2. Raíz cuadrada del error medio de predicción

Otro parámetro de evaluación del modelo son los errores de predicción, que resultan medidas útiles para seleccionar el número óptimo de factores a incluir en el modelo. Esto es debido a que los modelos son a menudo usados para predecir la concentración en futuras muestras desconocidas. Esto se realiza a través de los datos de validación del modelo, que se explicará en el siguiente ítem. La estadística usada para cuantificar el error en la predicción es la raíz cuadrada del error medio de predicción (RMSEP de *Root Mean Square Error Prediction*), que se puede observar en la ecuación 2.12.

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{\text{nmuestras}} \left(c_i - \hat{c}_i \right)^2}{\text{n muestras}}}$$

Ecuación 2.12.

donde c_i es la concentración observada de la i ésima muestra predicha, \hat{c}_i es la concentración predicha para esa muestra y n-muestras es el número de muestras predichas.

En función de lo expuesto en 2.2.5.1.1. y 2.2.5.1.2., existe una primera aproximación que permite seleccionar el número adecuado de componentes PLS a fin de obtener un modelo lo más ajustado posible. Esto se logra graficando la varianza no explicada, lo que es lo mismo a la raíz cuadrada de error medio (RMSEC) y la raíz cuadrada del error medio de predicción (RMSEP) en función del número de componentes PLS (figura 2.4.). Con el mismo, se puede elegir con mayor certeza, el número de componentes PLS óptimos necesarios para la obtención de un modelo adecuado [40].

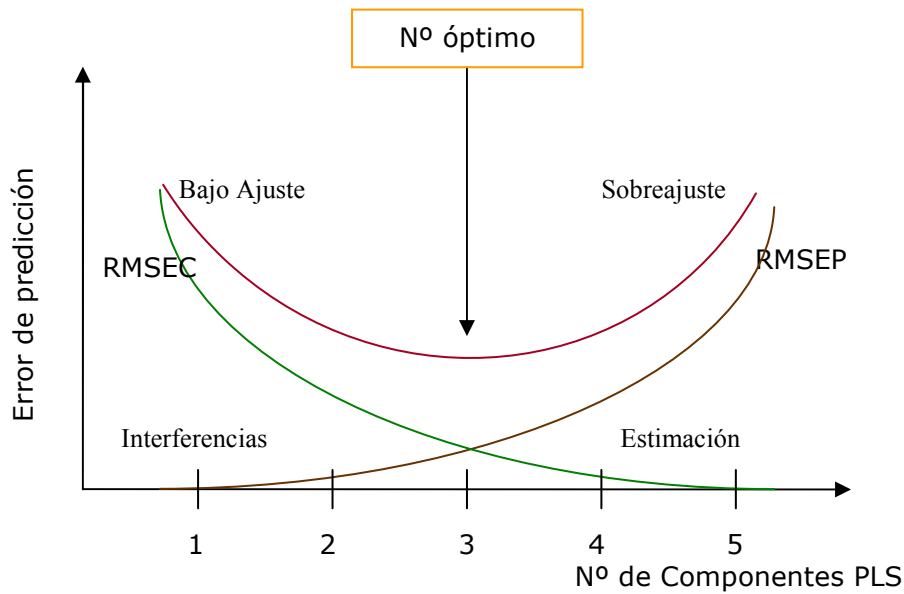


Figura 2.4. Estimación del número óptimo de componentes PLS para alcanzar el mejor ajuste del modelo.

2.2.5.1.3. Otros parámetros de evaluación del modelo

Existen, además de las nombradas recién, otras herramientas importantes en la evaluación del modelo. La Suma de Cuadrados del Error de Predicción (*PRESS* de *Prediction Error Sum of Squares*), es el error que surge como consecuencia del proceso de validación cruzada o *crossvalidación* del modelo, que se expondrá en el próximo ítem. El *PRESS* ofrece un criterio para definir el número de componentes PLS significativos del modelo. La ecuación 2.13 define matemáticamente al *PRESS*.

$$PRESS = \sum_i \sum_m [C_{im} - \hat{C}_{im}]^2$$

Ecuación 2.13

donde los subíndices i definen al i ésimo componente PLS, m corresponde al m ésimo analito de interés, C_{im} es la concentración experimental y \hat{C} corresponde a la concentración predicha del analito en los patrones de calibración.

Otro parámetro usado frecuentemente para evaluar al modelo es el factor Q^2 , que es la fracción de la variación total de la matriz R que puede ser explicada por cada uno de los componentes PLS y se calcula a partir del $PRESS$, como lo muestra la ecuación 2.14.

$$Q^2 = 1 - PRESS / SS$$

Ecuación 2.14

siendo SS la suma de cuadrados de las variables de R calculadas para cada uno de los componentes PLS.

Finalmente, un último parámetro útil para definir las variables más significativas de la matriz de respuestas R es el factor Importancia de Variables en la Proyección (VIP de *Variable Importance in the Proyection*). VIP calcula la influencia de cada k -ésima variable de la matriz de respuesta R sobre la matriz de concentraciones C y se calcula a través del factor Influencia de Variables (VIN de *Variable Influence*), tal como se muestra en la ecuación 2.15.

$$VIP_k = \sum [VIN]_k^2$$

Ecuación 2.15

2.2.5.2. Validación del modelo

Hay dos formas de generar un conjunto de validación para estimar los errores de predicción: validación interna (p.e. validación cruzada o crossvalidación) o validación externa (producir predicciones a partir de un conjunto separado de validación). La validación interna cruzada suele ser el método más usado por su sencillez.

Cuando se utiliza validación cruzada, la serie de datos de calibración original se divide en dos subconjuntos: calibración y predicción. Por ejemplo, la serie de datos original se divide a la mitad, usando la primera mitad de los datos para construir un modelo de calibración para predecir la segunda mitad. Los roles de los subconjuntos de datos son cambiados y se construye un nuevo modelo en donde la segunda mitad es usada para predecir la primera. De esta manera, cada una de las muestras en los datos es predicha una vez. Otro tipo de validación cruzada más común es el dejar *uno afuera*, donde los datos se dividen en subconjuntos conteniendo $n - 1$ muestras.

Las $n-1$ muestras son usadas para construir el modelo y la muestra afuera es usada para la predicción. El proceso se repite hasta que cada muestra es dejada afuera una vez. Para grandes series de datos, esto puede llevar importante cantidad de tiempo, por ello puede ser más práctico dejar afuera varios subconjuntos. Hay que tener en claro que existe un límite para cuantas y cuales muestras pueden ser dejadas afuera debido a que la serie de calibración debe contener un adecuado conjunto de muestras para construir el modelo a cada paso [25, 40, 42].

En la validación externa, se utilizan patrones de calibración distintos a los utilizados para la confección del modelo. Estos patrones se ingresan al modelo como si fueran muestras a analizar y se determinan con ellos la capacidad predictiva del modelo realizado, evaluando las diferencias de concentración que surgen entre la predicción del modelo y la concentración conocida de los analitos en los patrones de validación. Con ello es posible calcular el error asociado a fin de considerarlo para la estimación de las concentraciones en las muestras reales.

2.2.5.3. Predicción

La última etapa en la metodología PLS consiste en la predicción, en donde se cuantificarán las concentraciones de los analitos de interés en las diversas muestras incógnitas, a través de un modelo optimizado y validado.

Matemáticamente, la predicción consiste en conocer las matrices P, Q, E, F y G mientras que con las matrices T y U se calcula el vector b, que surgen a partir del modelo con matrices R y C conocidas y conformadas por los patrones de calibración.

A partir de una muestra de concentración desconocida para las especies de interés, se obtendrá un vector respuestas r_{desc} , con el cuál, a partir de la ecuación $r = t p + e$ (definida ahora en términos vectoriales para describir una sola muestra) se puede calcular el vector t_{desc} . Como el vector b y g ya están dados por el modelo, con la expresión $u = bt + g$, con t_{desc} se calcula u_{desc} . Finalmente con la ecuación $c = uq + f$ (vectorial) y conocidos todos los miembros de la misma, es posible conocer el vector c_{desc} , que resulta ser un vector fila cuyos componentes corresponden las concentraciones de los analitos de interés previstos en el modelo [41].

2.2.6. Paquetes Estadísticos para el Empleo de la Metodología PLS

El análisis a través de metodología PLS requerirá la utilización de programas específicos para tal fin, debido a la complejidad de los cálculos que se deben realizar. Además tales programas pueden ser lo suficientemente accesibles como para facilitar la tarea al operador. Sin embargo, el uso de herramientas multivariadas requerirá del usuario, en general, un conocimiento lo más acabado posible de los fundamentos matemáticos necesarios, como así también del sistema que desee estudiar y del programa estadístico del que dispone, a fin de no obtener resultados incorrectos o falsos. Esto resulta algo muy importante ya que no es posible trabajar con estos programas sin una base sustentable, ya que los resultados que se obtengan pueden engañar al analista y emitirse resultados muy alejados de la realidad.

Existen en el mercado, un número importante de programas poderosos de computación que permiten el análisis multivariado de los datos analíticos. Además hay una oferta muy grande de programas estadísticos multivariados para diferentes propósitos. Esto obliga al analista químico que los usará, conocer cabalmente cada uno de ellos para que le permita realizar la selección del mejor programa sin gastar dinero en paquetes que le pudieran prestar poca utilidad.

El programa utilizado Unscrambler 6.11 [45] está diseñado con una interfase del usuario accesible que permite, entre otros aspectos, importar las matrices de datos desde cualquier planilla de cálculo, obtener gráficas espectrales (completas o acotadas), gráficos específicos referidos a los modelos obtenidos, que ayudan a interpretar y a optimizar los mismos en poco tiempo (si no existen otro tipo de fuente de error adicional). Esto facilita enormemente la tarea del químico analítico para el tratamiento de datos y le evita considerables pérdidas de tiempo en la transcripción de los mismos, sumado ha ello la posibilidad de error que es posible cometerse durante el proceso por parte del operador, habida cuenta de que cada patrón o muestra puede estar constituido por cientos de datos.

El Unscrambler 6.11 permite realizar el tratamiento de datos analíticos bajo diversas formas de herramientas multivariadas. Las utilizadas en este trabajo son PLS-1 y PLS-2, pero también posee tratamiento de datos por medio de PCR y PCA (este último importante para clasificación por grupos a través de ciertas propiedades características, como así también permite encontrar propiedades ocultas relevantes que puedan afectar enormemente el análisis).

Otra característica interesante del sistema Unscrambler es la inclusión de un manual del usuario [46] sumamente accesible, con una gran diversidad de ejemplos para cada caso particular, lo cual facilita la interpretación y entrenamiento del analista. Por otra parte, el fabricante incluye un programa tutorial que está basado en los ejemplos antes mencionados y con la misma interfase del usuario, que permite realizar un seguimiento más exhaustivo de todo el procedimiento y un mejor entendimiento del mismo paso por paso, lo cual redundará en un menor tiempo de aprendizaje por parte del analista neófito.

CAPÍTULO 3. MATERIALES Y MÉTODOS

3.1. REACTIVOS

3.1.1. Soluciones Estándar

3.1.1.1. Solución de Sacarina Sódica.

Se pesó 1,0000 g de sacarina sódica p.a. (Merck), se trasvasó a un matraz de 1000 mL y se adicionó agua desionizada hasta enrase.

3.1.1.2. Solución de Aspartamo

Se pesó 1,0000 g de aspartamo p.a. (Parapharm), se trasvasó a un matraz de 1000 mL y se adicionó agua desionizada hasta enrase.

3.1.1.3. Solución de Ácido Salicílico

Se pesó 1,0000 g de ácido salicílico p.a. (Merck), se trasvasó a un matraz de 1000 mL y se adicionó agua desionizada hasta enrase.

3.1.1.4. Solución Reguladora pH 2.

Se preparó esta solución reguladora utilizando la tabla de Clark, a partir de soluciones estándar de HCl y KCl [47].

3.1.1.4. Agua Deionizada

Se obtuvo a partir de un sistema de ultrapurificación de agua Barnstead Easy Pure (USA) con una resistividad de 18 MΩ cm.

3.2 INSTRUMENTAL

3.2.1 Medidas Absorciométricas

Las medidas absorciométricas y los datos espectrales correspondientes, fueron realizados dentro del rango visible, en un espectrofotómetro marca Ocean Optics modelo CHEMUSB4 con sistema de detección por arreglo de diodo lineal y de configuración haz simple. Las celdas utilizadas en todos los casos fueron de cuarzo y con un camino óptico de 10 mm. Los datos fueron transferidos y procesados en el programa Unscrambler 6.11. Todos los blancos de lectura cero en escala del instrumento fueron realizados con agua deionizada adicionada con solución reguladora pH 2.

3.2.2. Medidas y Ajuste de pH

Para las medidas de pH se utilizó un pH-metro HORIBA F42 de electrodo combinado de vidrio, con electrodo de referencia interna de Ag-AgCl.

CAPÍTULO 4. RESULTADOS Y DISCUSIÓN

4.1. MODELO DE CALIBRACIÓN OBTENIDO A TRAVÉS DE REGRESIÓN POR MÍNIMOS CUADRADOS PARCIALES (PLS)

El modelo de calibración fue obtenido utilizando un diseño de experimento factorial completo, adoptándose el criterio de cuatro niveles de concentración y dos analitos, obteniéndose así un total de $2^4 = 16$ patrones de calibración.

Por otra parte, se utilizó un patrón interno de ácido salicílico, el cual fue agregado tanto a los patrones de calibración como a las muestras analizadas. Este patrón fue empleado para evaluar el grado de recuperación en muestras de edulcorantes desconocidas, tal que si la concentración predicha por el modelo para este analito es igual o similar a la agregada, no hay influencias de matriz de la muestra sobre el modelo PLS obtenido.

La figura 4.1. muestra el grado de solapamiento espectral observado para sacarina, aspartamo y ácido salicílico, este último empleado como patrón interno.

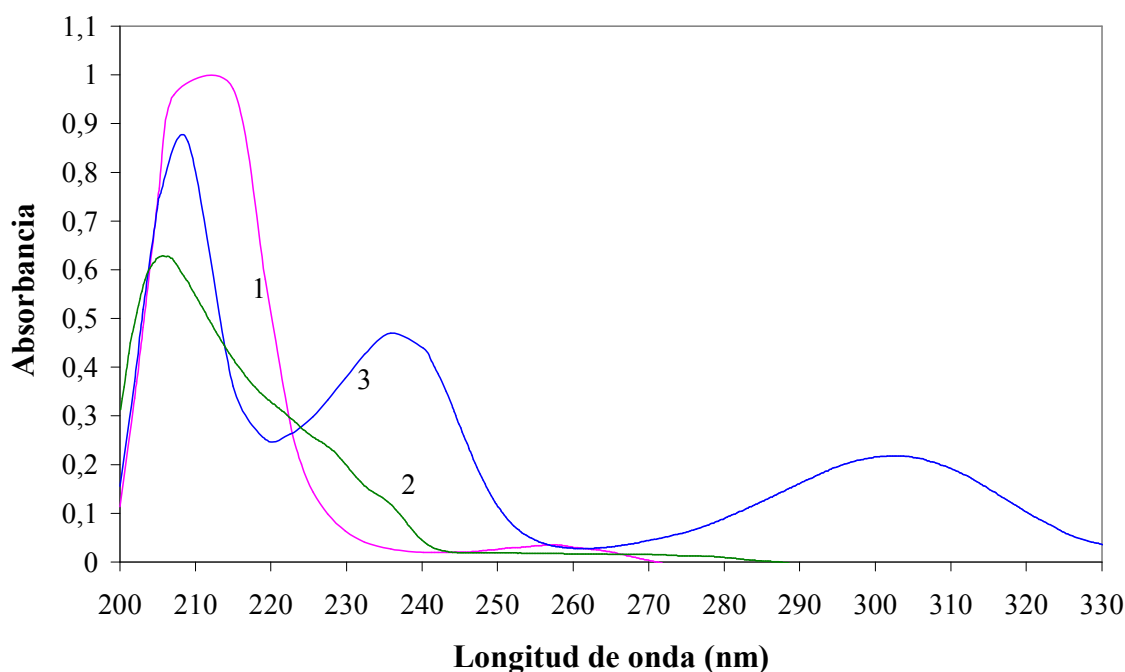


Figura 4.1. Curvas espectrales de (1) aspartamo (0.005 g L^{-1}), (2) sacarina (0.005 g L^{-1}), (3) estándar interno (0.008 g L^{-1}).

La tabla 4.1. muestra el conjunto de calibración utilizado para obtener el modelo PLS con 16 patrones, correspondientes a 4 niveles de concentración y dos analitos. El ácido salicílico fue agregado en forma aleatoria para evaluar el grado de recuperación, pero no intervino en el diseño de experimento por no ser analito de interés, sino solamente ser empleado para evaluar la adecuada predicción del modelo en las muestras.

Tabla 4.1. Conjunto de calibración utilizado para obtener el modelo PLS con 16 patrones

	Aspartamo ^a	Sacarina ^a	Patrón Interno ^a
Patrón 1	0,000	0,000	0,000
Patrón 2	0,001	0,000	0,002
Patrón 3	0,003	0,000	0,008
Patrón 4	0,005	0,000	0,000
Patrón 5	0,000	0,010	0,004
Patrón 6	0,001	0,010	0,002
Patrón 7	0,003	0,010	0,008
Patrón 8	0,005	0,010	0,000
Patrón 9	0,000	0,030	0,008
Patrón 10	0,001	0,030	0,002
Patrón 11	0,003	0,030	0,004
Patrón 12	0,005	0,030	0,004
Patrón 13	0,000	0,050	0,008
Patrón 14	0,001	0,050	0,004
Patrón 15	0,003	0,050	0,002
Patrón 16	0,005	0,050	0,000

^a Concentración expresada en g.L

4.2. EVALUACIÓN DEL MODELO PLS

Utilizando el diseño de experimento mencionado en 4.1, se obtuvo un modelo PLS-2 cuyas propiedades fueron analizadas para evaluar la capacidad predictiva del mismo. Se utilizaron las longitudes de onda comprendidas entre los 200 a 330 nm con lecturas obtenidas cada 0,5 nm. En figura 4.2. se observan los espectros obtenidos para los 16 patrones de calibración, con cuyos datos absorciométricos fue posible obtener el modelo PLS.

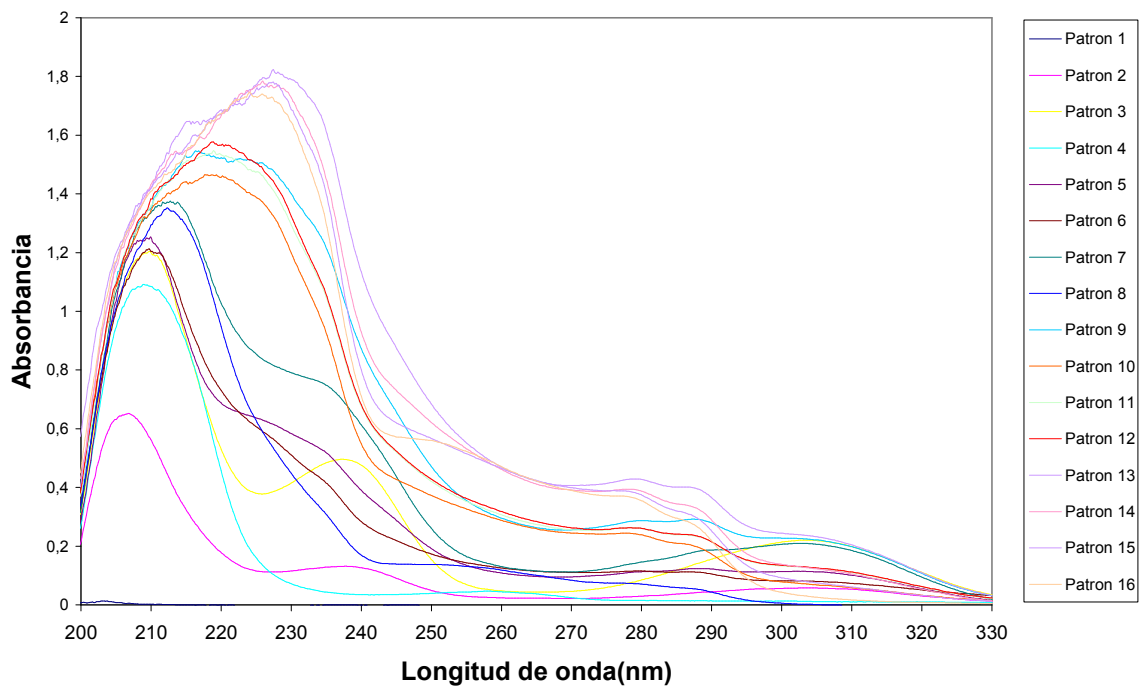


Figura 4.2. Curvas espectrales de los 16 patrones de calibración.

El modelo fue validado, en principio, internamente a partir de una validación interna cruzada. Posteriormente se realizó una segunda validación en las muestras reales a través del empleo del estándar interno. A través de la validación cruzada, fue posible obtener los errores del modelo, a través de la raíz cuadrada media del error de calibración (RMSEC) y la raíz cuadrada media del error de predicción (RMSEP). Para ambos casos se busca obtener un mínimo, el cual indica el número óptimo de componentes principales a utilizar en el modelo. La figura 4.3. muestra las curvas de RMSEC y RMSEP obtenidas para el modelo PLS empleado, el cual indica que el número óptimo para el modelo propuesto fue de 4 componentes principales.

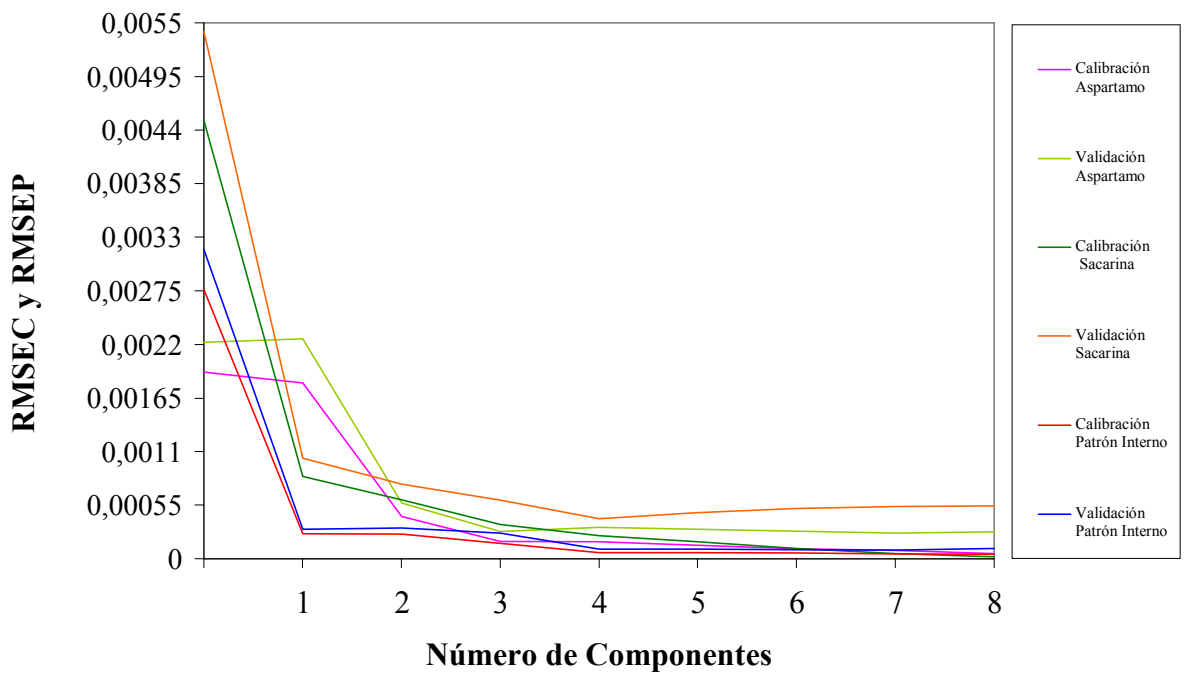


Figura 4.3. Curvas de RMSEC y RMSEP obtenidas para el modelo PLS empleado

La figura 4.4. muestra el porcentaje de varianza explicada, cuyo gráfico indica la cantidad de información de las variables originales acumulada en los componentes principales. Como se puede apreciar, se muestran las curvas para sacarina, aspartamo y ácido salicílico, en la que se pueden observar que fueron necesarios cuatro componentes principales para explicar casi el 98,81 % de la información original.

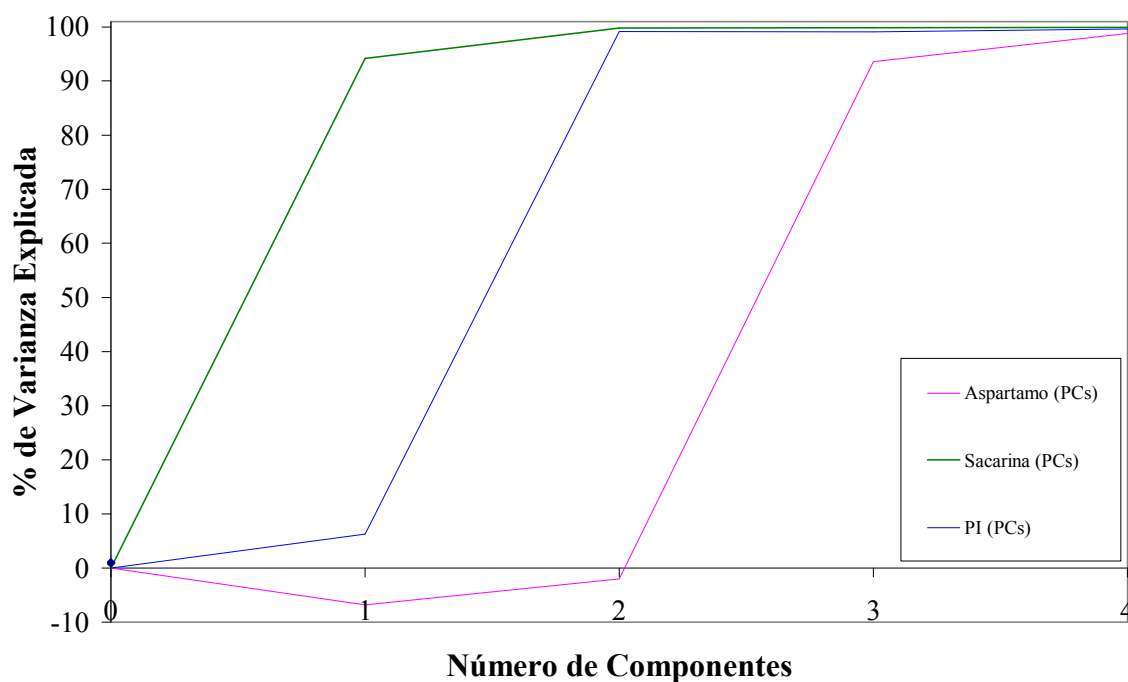


Figura 4.4. Porcentaje de Varianza explicada en función del número de componentes PLS.

4.3. EVALUACIÓN DE SACARINA, ASPARTAMO Y PATRÓN INTERNO EN EL MODELO DE CALIBRACIÓN

Una vez halladas las condiciones óptimas del modelo, a través de una adecuada selección del número de componentes principales, se realizaron las curvas de calibración de concentraciones observadas-predichas para los dos analitos de interés y el patrón interno de ácido salicílico. En una situación ideal, dichas gráficas deben poseer una pendiente igual a 1, una ordenada al origen igual a cero y un valor de coeficiente de regresión cercano a 1. Las figuras 4.5., 4.6. y 4.7. muestran los gráficos de concentraciones observadas-predichas para aspartamo, sacarina y patrón interno de ácido salicílico respectivamente. Se observa que las gráficas se acercan a los parámetros ideales para este tipo de calibración. En todos los casos, las concentraciones están expresadas en mg L^{-1} .

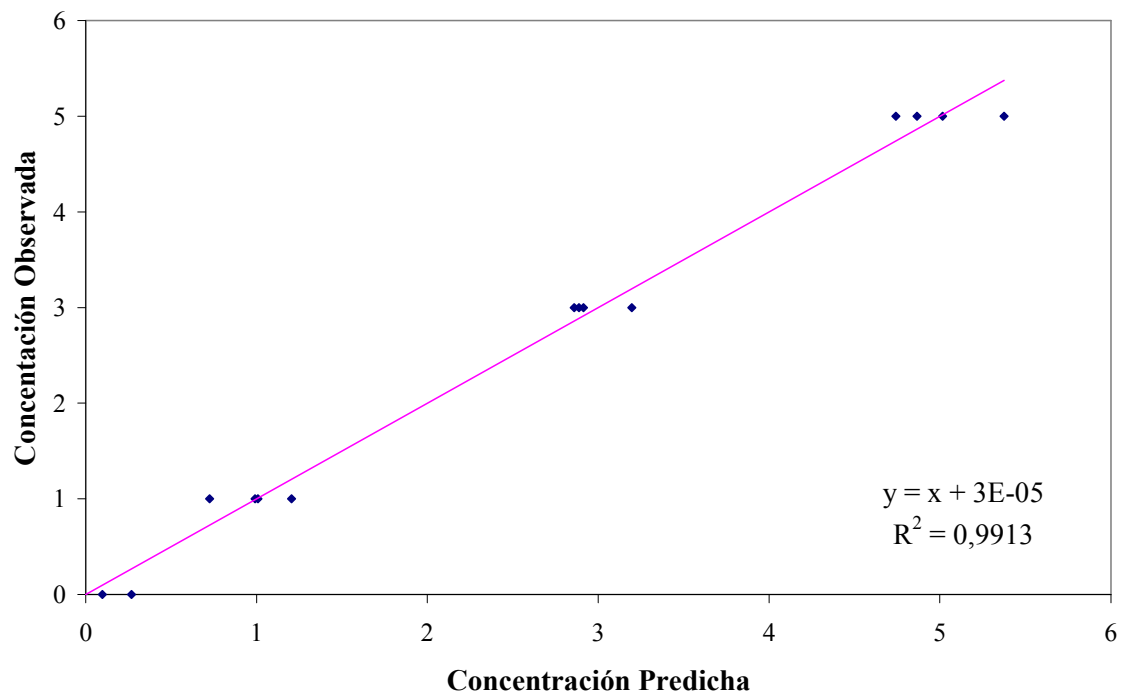


Figura 4.5. Concentración Observada- Predicha de Aspartamo (en mg L^{-1})

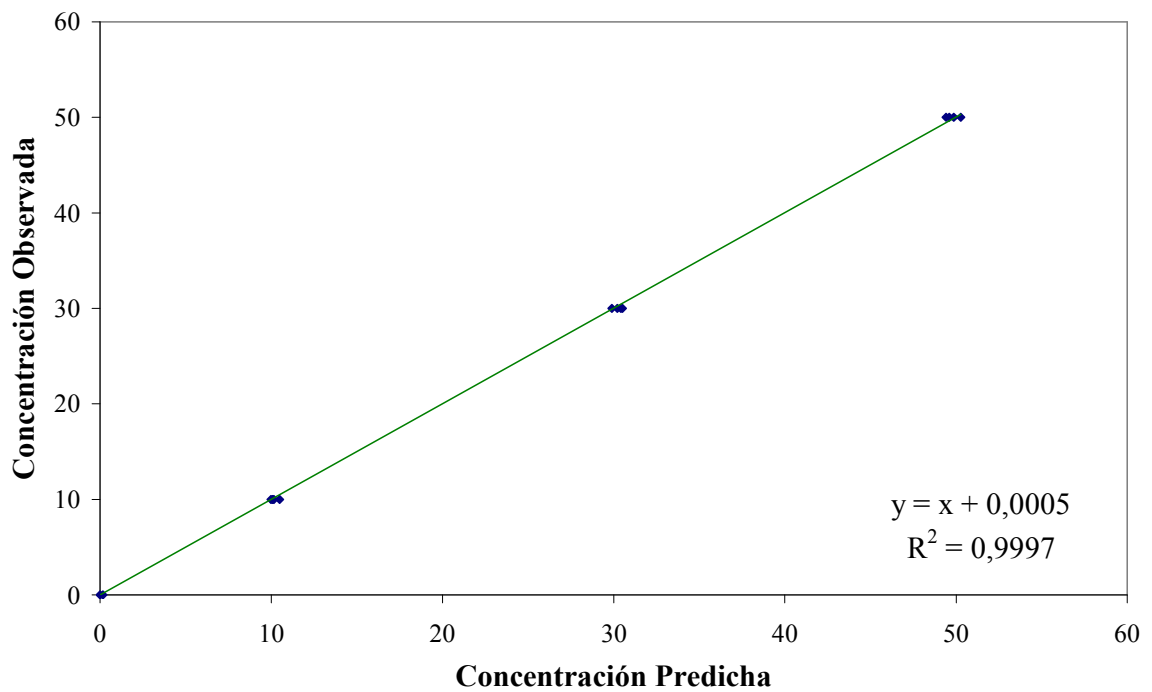


Figura 4. 6. Concentración Observada- Predicha de Sacarina (en mg L^{-1})

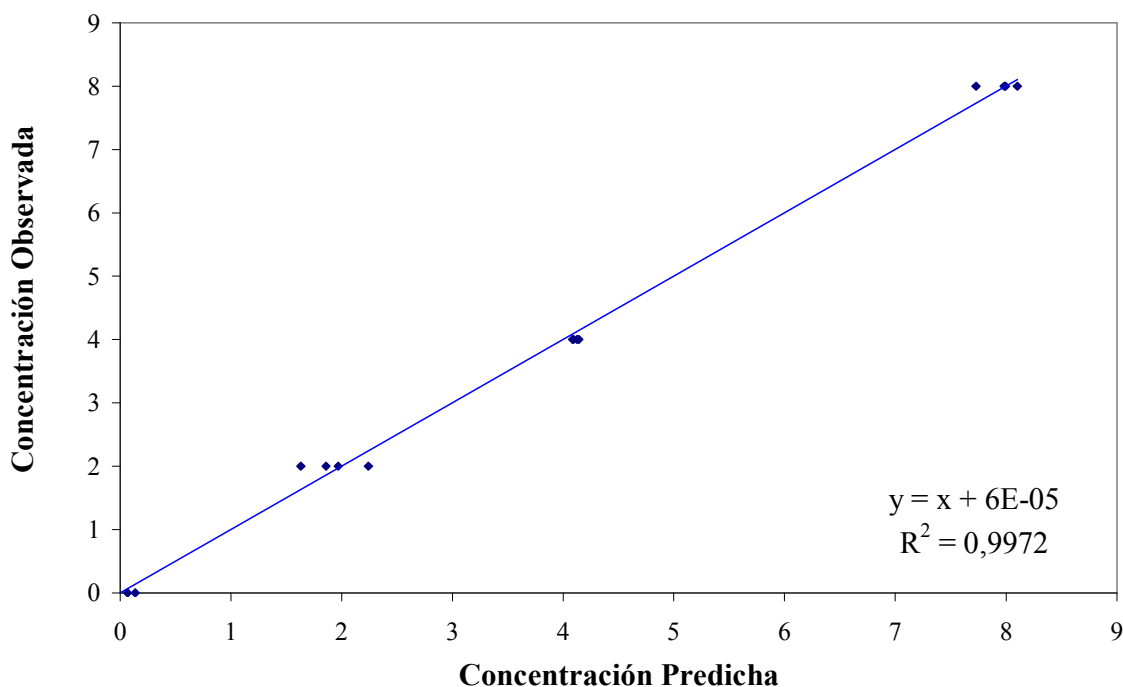


Figura 4.7. Concentración Observada- Predicha de Patrón Interno de Ácido salicílico (en mg L⁻¹)

4.4. EVALUACIÓN DE EFECTOS DE MATRIZ DE LA MUESTRA SOBRE EL MODELO PLS.

Se utilizaron 5 muestras de edulcorantes comerciales artificiales, de las cuales 3 fueron líquidas (1: Semblé; 2: Chuker y 3: Hileret) y 2 sólidas en tabletas (4: Sucaryl y 5: Chuker). Las muestras líquidas fueron diluidas de tal manera de que la concentración declarada en el envase de ambos analitos se encuentre dentro del rango de concentraciones de la matriz de calibración. Para el caso de muestras sólidas, las mismas fueron pesadas con un error de $\pm 0,0001$ g, fueron disueltas y diluidas tal que la concentración de ambos analitos se encontrara dentro del rango de la matriz de calibración. En todos los casos, previo al enrase se adicionaron cantidades conocidas de ácido salicílico, a fin de evaluar el grado de ajuste del modelo sobre las muestras analizadas. La Tabla 4.2. muestra los resultados obtenidos para evaluar el comportamiento del modelo en las 5 muestras estudiadas a través del uso de patrón interno. En la misma se observa que la predicción del modelo sobre el patrón interno en las muestras reales (resultados obtenidos a partir de 5 réplicas para cada una) tuvo un adecuado ajuste, observándose una recuperación media de 101,2 %, con un rango comprendido entre 97,5 a 110,0 %.

Tabla 4. 2. Comportamiento del patrón interno utilizado en muestras reales de edulcorante.

Muestras	Patrón Interno			
	Agregado ^a	Predicho ^{a,b}	Recuperación (%) ^b	% Er ^c
Liquida 1	0,0040	0,0039 ± 0,0002	97,5 ± 5,2	-2,5
Liquida 2	0,0010	0,0011 ± 0,0003	110,0 ± 8,1	10,0
Liquida 3	0,0080	0,0079 ± 0,0002	98,8 ± 2,6	-1,3
Tableta 1	0,0060	0,0059 ± 0,0002	98,3 ± 3,8	-1,7
Tableta 2	0,0080	0,0081 ± 0,0001	101,3 ± 3,0	1,3

^a Concentración expresada en g L⁻¹

^b Promedio ± Desviación Estándar de cinco determinaciones

^c Porcentaje de Error Relativo de Recuperación

4.5. CUANTIFICACIÓN DE SACARINA Y ASPARTAMO EN MUESTRAS REALES DE EDULCORANTES

Una vez evaluada la capacidad predictiva del modelo a través del uso de estándar interno, se analizaron las muestras citadas en el punto 4.4., las cuales estaban adicionadas con el patrón interno.

La tabla 4.3. muestra los resultados para la cuantificación de sacarina y aspartamo en las cinco muestras reales analizadas, los que fueron obtenidos a partir de la lectura de 5 réplicas por muestra. Se observan las concentraciones predichas por el modelo para cada una de las muestras y los valores nominales (declarados en etiqueta del producto). Los errores obtenidos fueron menores al 12 % en todos los casos.

Tabla 4.3. Concentraciones predichas en muestras reales utilizando PLSR-2.

Muestra	Aspartamo		Sacarina	
	PLSR-2	Valor nominal	PLSR-2	Valor nominal
Liquida 1 ^a	1,48 ± 0,10	1,5	27,44 ± 0,18	27,0
Liquida 2 ^a	1,47 ± 0,14	1,5	26,76 ± 0,15	27,0
Liquida 3 ^a	4,98 ± 0,70	5,0	26,82 ± 0,26	27,0
Tabletas 1 ^b	37,65 ± 0,36	37,0	99,21 ± 0,35	100,0
Tabletas 2 ^b	36,92 ± 0,24	37,0	96,77 ± 0,39	100,0

^a Concentraciones de edulcorantes no diluidos expresados en g L⁻¹

^b Concentraciones de edulcorantes en tabletas expresados en mg g⁻¹

CAPÍTULO 5. CONCLUSIONES

5. 1. DEL PRETRATAMIENTO DE MUESTRA.

El método multivariado propuesto presenta una ventaja importante respecto a la posibilidad de realizar las determinaciones analíticas sin la necesidad de utilizar métodos separativos previos ni ningún tipo de pretratamiento de la muestra. Esto es importante debido a que las muestras reales presentan una composición de especies que pueden actuar como matriz de interferencias.

5. 2. DE LOS COSTOS INSTRUMENTALES.

En bibliografía existen un gran número de trabajos relacionados a la determinación simultánea de sacarina y aspartamo en muestras comerciales de edulcorantes, jugos, etc. Sin embargo, estas metodologías publicadas incluyen instrumental de muy elevado costo, lo cual lo hace poco accesible a pequeños laboratorios de control de calidad o laboratorios de bajos recursos.

5. 3. DE LAS VENTAJAS MULTIVARIADAS.

Los métodos actuales de análisis implican el uso de pocas variables instrumentales. Por ello resulta de interés la exploración de nuevos métodos en donde se puedan emplear un gran número de variables para el análisis simultáneo. Este es el caso de los multivariados que utilizan una gran cantidad de datos instrumentales y en el caso de esta tesis, consistieron en obtener los valores de absorbancia de espectros completos, generándose así matrices de respuestas de gran tamaño, que no serían factibles de analizar con metodologías de cálculo convencional.

5. 4. DE LA APLICABILIDAD DEL MÉTODO

Debido a su bajo costo de implementación, sumado a los bajos errores obtenidos en comparación a los valores de referencia del fabricante, la metodología multivariada propuesta es adecuada para laboratorios de control de calidad, laboratorios de rutina y de bajos recursos presupuestarios, para la cuantificación rápida y confiable de sacarina y aspartamo en edulcorantes comerciales.

CAPÍTULO 6. BIBLIOGRAFÍA

1. J.E.C.F.A. (Joint Expert Committee on Food Additives)
2. C.A.A. (Código Alimentario Argentino). (Actualizado)
3. Goette, T. (2000) “Suma Bromatológica, Tratado de Bromatología. Teórico – Práctico”. Training Tec – IPSA, 1º Edición.
4. Sistema internacional de numeración para los aditivos alimentarios.htm
5. International Sweeteners Association, Brussels, Belgium, 2004; <http://www.isabru.or/>
6. Mitchell, M., Pearson, R.L., Saccharin, in R.C. Gelardi (Ed.), (1991) “Alternative Sweetener”, Marcel Dekker Inc, New York, , pp. 127 -156.
7. U.S. Department of Health and Human Services, Public Health Service. National Institutes of Health. Ntp Technical Report on the Toxicity Studies of Aspartame (Case No. 22839-47-0); U.S. GPO: Washington, DC, 2003.
8. Wasik, A.; McCourt, J.; Buchgraber, M. (2007) “Simultaneous determination of nine intense sweeteners in foodstuffs by high performance liquid chromatography”. *J. Chromatogr. A*, 1157, 187-96.
9. Wu, P. W.; Cheng, C. C.; Chou, S. S. (1995) “Determination of acesulfame-K, saccharin, aspartame, dulcin and cyclamate in beverages”. *J. Chin. Agric. Chem. Soc.*, 33, 37-50.
10. Zhu, Y.; Guo, Y.; Ye, M.; James, F.S. (2005) “Separation and simultaneous determination of four artificial sweeteners in food and beverages by ion chromatography”. *J. Chromatogr. A*, 1085, 143-146.
11. Qu, F.; Qi, Z. H.; Liu, K. N.; Mou, S. F. (1999) “Determination of aspartame by ion chromatography with electrochemical integrated amperometric detection”. *J. Chromatogr. A*, 850, 277-281.
12. Herrmannova, M.; Krivankova, L.; Bartos, M. (2006) “Direct simultaneous determination of eight sweeteners in foods by capillary isotachopheresis”. *J. Sep. Sci.*, 29, 1132 -1137.
13. Frazier, R.A.; Inns, E. L.; Dossi, N. ; Ames, J.M.; Nursten, H.E. (2000) “Development of a capillary electrophoresis method for the simultaneous analysis of artificial sweeteners, preservatives and colours in soft drinks”. *J. Chromatogr. A*, 876, 213-220.

14. Armenta, S.; Garrigues, S.; De La Guardia, M. (2004) "FTIR determination of aspartame and acesulfame-K in table top sweeteners". *J. Agric. Food Chem.*, *52*, 7798-7803.
15. Khurana, H.K.; Cho, I.K.; Shim, J.Y.; Li, Q.X.; Jun, S. (2008) "Application of multibounce attenuated total reflectance fourier transform infrared spectroscopy and chemometrics for determination of aspartame in soft drinks". *J. Agric. Food Chem.*, *56*, 778-783.
16. Armenta, S.; Garrigues, S.; De La Guardia, M. (2004) "Sweeteners determination in table top formulations using FT-Raman spectrometry and chemometric analysis". *Anal. Chim. Acta*, *521*, 149-155.
17. Nikolelis, D. P.; Pantoulis S. (2000) "A minisensor for the rapid screening of acesulfame-K, cyclamate and saccharin based on surface-stabilized bilayer lipid membranes". *Electroanalysis*, *12*, 786-790.
18. Fatibello-Filho, O.; Marcolino-Junior, LH; Pereira, AV. (1999) "Solid-phase reactor with copper(II) phosphate for flow-injection spectrophotometric determination of aspartame in tabletop sweeteners". *Anal. Chim. Acta*, *384*, 167-174.
19. Odaci, D; Timur, S.; Telefoncu, A. (2004) "Carboxyl esterase-alcohol oxidase based biosensor for the aspartame determination". *Food Chem.*, *84*, 493-496.
20. Fatibello-Filho, O.; Aniceto, C. (1997) "Potentiometric determination of saccharin in dietary products using a coated-carbon rod ion selective electrode". *Anal. Lett.*, *30*, 1653-1666.
21. Filho, J.C.; Oppermann Santini, A.; Nasser, A.L.M.; Pezza, H.R.; de Oliveira, J.E.; Melios, C.B.; Pezza, L. (2003) "Potentiometric determination of saccharin in commercial artificial sweeteners using a silver electrode". *Food Chem.*, *83*, 297-301.
22. Pena, R. M.; Lima, J. L. F. C.; Saraiva, M.; Lucia, M. F. S. (2004) "Sequential injection analysis-based flow system for the enzymatic determination of aspartame". *Anal. Chim. Acta*, *514*, 37-43.
23. Vallvey, C. L. F.; Valencia, M. C. E.; Nicolás, A.; García-Jiménez, J. F. (2006) "Resolution of an intense sweetener mixture by use of a flow injection sensor with on-line solid-phase extraction. Application to saccharin and aspartame in sweets and drinks". *Anal. Bioanal. Chem.*, *385*, 385-391
24. Wold, S. and Eriksson, L. in: H. Van der Waterbeemd Ed. (1995). "Chemometric Method in Molecular Design". VCH Weinheim (Alemania).

25. Cela, R. (1994). "Quimiometría Práctica". Publicaciones Universidad de Santiago de Compostela (España).
26. Mássart, D.; Vandeginste, B.; Buydens, L.; de Jong, S.; Lewy, P. and Smeyers-Verbeke, J. (1997). "Handbook of Chemometrics and Qualimetrics: Part A and B". Elsevier (Alemania).
27. Camiña, J.; Ferreyra, R.; Luco, J. and E. Marchevsky. (2001). *Journal of Analytical Chemistry* 56 (1) 26.
28. Ferreyra, R.; Camiña, J.; Marchevsky, E. and J. Luco. (2000) *Fressenius Journal of Analytical Chemistry* 368 (6) 32.
29. Camiña, J. "Resolución de Elementos de Tierras Raras y Otros Metales de Interés en Superconductores por la Aplicación de Métodos Quimiométricos". Tesis de Maestría. UNSL. 1999.
30. Preisler, J.; Jankar, L. and Sommer, M. (1993). *Collect. Czech. Chem. Commun* 58, 1495.
31. Bautista, R.; Jimenez, A.; Jimenez, F. and Arias, J. (1996). *Talanta* 43, 421.
32. Jankar, L.; Preisler, J. and Sommer, M. (1993). *Collect. Czech. Chem. Commun* 58, 1509.
33. Perutka, J.; Havel, J. and Jancar, L. (1994). *Collect. Czech. Chem. Commun* 59, 1951.
34. Navalón, A., Blanc, R.; Del Olmo, M. and Vilchez, J. (1999). *Talanta* 48, 469.
35. Ramos Martos, N.; Molina Díaz, A.; Navalón, A.; De Orbe Payá, I. and Capitán Vallbey, L. (2000). *Journal of Pharmaceutical and Biomedical Analysis* 23, 837.
36. Coscione, A.; De Andrade, J.; Poppi, R.; Mello, C.; Van Raij, B. and Ferreira de Abreu, M. (2000). *Analytica Chimica Acta* 423, 31.
37. Kubán, V.; and Gladilovich, D. (1988) *Coll Czech. Chem. Commun*, 53, 1664.
38. Wentzell, P.; Andrews, D. and Kowalski, B. (1997) *Anal. Chem*, 69 (13), 2299.
39. Martents, H. and Naes, T. (1989). "Multivariate Calibration". John Wiley and Sons (Nueva York).
40. Beebe, K.; Pell, R. and Seasholtz, M. (1998). "Chemometrics, a Practical Guide". John Wiley and Sons (New York).
41. Beebe, K and Kowalsky, B. (1987) *Analytical Chemistry* 59 (17) 1007.
42. Kramer, R. (1998). "Chemometric Techniques for Cuantitative Analysis". Marcel Dekker (New York).

43. Esbensen, K.; Schönkopf, S. and Midgaard, T. (1997). "The Unscrambler. Multivariate Analysis in Practice". CAMO-AS (Noruega).
44. Otto, M. (1999). "Chemometrics". Wiley-VCH (Alemania).
45. The Unscrambler v 6.0 by CAMO-AS, Trondheim (Noruega).
46. The Unscrambler 6. User's Guide. CAMO-AS, Trondheim (Noruega).
47. Meites, L. (1982) "Handbook of analytical chemistry". Mc Graw-Hill (New York).