



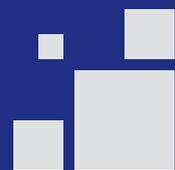
CENATAV
Centro de Aplicaciones de
Tecnologías de Avanzada
MINISTERIO DE LA INDUSTRIA BÁSICA

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

SERIE AZUL

RNPS No. 2142
ISSN Solicitado

7ma. No. 21812 e/218 y 222,
Rpto. Siboney, Playa;
Ciudad de La Habana.
Cuba. C.P. 12200
www.cenatav.co.cu





CENATAV

Centro de Aplicaciones de
Tecnologías de Avanzada
MINISTERIO DE LA INDUSTRIA BÁSICA

REPORTE TÉCNICO

Reconocimiento de Patrones

SERIE AZUL

Estado del Arte del Reconocimiento de Patrones en la Quimiometría

Isneri Talavera Bustamante
Jorge Luis Rodríguez Hierrezuelo

RT_002

Julio 2008

RNPS No 2142

ISSN Solicitado

7ma. No. 21812 e/218 y 222,

Rpto. Siboney, Playa;

Ciudad de La Habana.

Cuba. C.P. 12200

www.cenatav.co.cu



Estado del Arte del Reconocimiento de Patrones en la Quimiometría

Dra. Isneri Talavera Bustamante, Ing. Jorge Luis Rodriguez Hierrezuelo

Centro de Aplicaciones de Tecnologías de Avanzada, 7a #21812 e/ 218 y 222, Siboney, Playa, Habana, Cuba
italavera@cenatav.co.cu

RT_002 CENATAV

Fecha del camera ready: 12 de noviembre de 2005

Resumen: Este reporte de investigación presenta en forma detallada un recuento del quehacer en el área de la quimiometría, con especial énfasis en las técnicas de reconocimiento de patrones. Se comienza con una explicación acerca de qué es la quimiometría, como surgió hace 30 años y su rápida evolución hasta nuestros días. Se presenta un detallado recuento histórico de la especialidad, y se aborda a continuación las principales áreas de trabajo de la quimiometría y se comentan individualmente, incluyéndose un ejemplo numérico para el caso del análisis jerárquico de agrupamientos. Se ofrece una idea concisa acerca de los métodos de reconocimiento de patrones tales como: PCA, HCA, LDA, SIMCA, KNN, LMM, Redes Neuronales y otros. El reporte incluye además información sobre los diferentes grupos de quimiometría en diferentes países y sus principales áreas de investigación. Un conjunto amplio de libros publicados y una extensa relación de referencias bibliográficas relacionadas con los métodos complementa la información. A modo de conclusión el presente reporte ofrece a la comunidad de quimiometristas un resumen crítico del estado actual de la quimiometría y sus tendencias futuras de desarrollo.

Palabras clave: quimiometría; reconocimiento de patrones.

Abstract: The investigation report presents a detailed overview of the area of chemometrics with special emphasis on pattern recognition techniques. It starts explaining what is chemometrics, how it was born about 30 years ago and its quickly evolution up to our days. It contains a detailed historical account. Next, the main areas of chemometrics are presented and each of them is commented individually, including a numerical example on Hierarchical Cluster Analysis. A concise idea about all pattern recognition methods such as PCA, HCA, LDA, SIMCA, KNN, LMM, Neural Networks, and others, is given. The report also includes the information of several chemometrics groups in different countries around the world and their main focus researches. The complete set of books published in this area is given, and also an extensive list of references regarding the different methods. As a conclusion, this report offers to the chemometric community a critical review of the state of the art of chemometrics and its trends for the future.

Keyword: chemometrics; pattern recognition.

1 Introducción

El presente reporte es el primero de una serie, que tiene como objetivo principal brindar un producto informativo valioso referido al tema de la Quimiometría, que permita introducir y motivar a una masa crítica de químicos cubanos en el estudio, aplicación y desarrollo de esta importante disciplina química.

Este primer trabajo recoge las principales características de la Quimiometría. Se detalla su finalidad, su posición entre las ciencias experimentales, sus principales áreas de trabajo, desarrollando aquella que aborda las investigaciones y aplicaciones actuales más relevantes relacionadas con el Reconocimiento de Patrones y el Análisis Exploratorio de Datos, por ser precisamente en ésta área, donde se desarrolla el principal accionar de los autores. Se brinda además un resumen de los principales grupos de trabajo y personalidades en el campo de la Quimiometría a nivel mundial, así como las principales publicaciones (revistas, páginas Web, libros etc.) relacionadas con el tema, que facilite a los lectores la continuación y profundización de los tópicos de interés.

De interés resulta el epígrafe referido a análisis de los software más utilizados en la especialidad, que tienen la característica de incorporar herramientas estadísticas específicas fruto de las investigaciones en Quimiometría, y poco extendidas entre las otras especialidades de la Estadística aplicada. Se concluye con un análisis dirigido a orientar a nuestros quimiometristas hacia donde se encuentran los principales retos investigativos tanto teóricos como aplicados en el área del Reconocimiento de Patrones en la Quimiometría en el contexto mundial, y su correspondencia con la solución de los problemas nacionales.

Para la elaboración de este informe fueron consultadas un variado grupo de fuentes bibliográficas entre las que se encuentran: Revisión de la base de datos de la Chemical Abstract, estudio y análisis de más de 200 artículos correspondientes a la temática, análisis de 4 recopilaciones de resultados sobre quimiometría publicada en la Analytical Chemistry (1995-2004), consulta de más de 50 sitios y páginas Web de la especialidad y estudio de 3 libros de textos.

1.2 Definición y evolución histórica de la quimiometría

La automatización y computarización de los laboratorios ha llevado consigo diversas consecuencias. Una de ellas es la rápida adquisición de gran cantidad de datos. Ahora bien, sabemos que la posesión de dichos datos dista, muchas veces, de proporcionar respuestas adecuadas. Obtener de datos no es sinónimo de poseer información; debemos interpretarlos y colocarlos en el contexto adecuado para convertirlos en información útil para el usuario. La **Quimiometría** es la disciplina que tiene esta finalidad. La palabra quimiometría, inventada hace aproximadamente treinta años, quiere resumir el concepto que engloba la medida en química.

D.L Massart. Uno de los fundadores de la Quimiometría definió la misma [1], como “la disciplina química que utiliza la Matemática, la Estadística y la lógica formal (a) para diseñar o seleccionar procedimientos experimentales óptimos; (b) proporcionar la máxima información química relevante a partir del análisis de datos químicos; (c) y obtener conocimiento de sistemas químicos.

Por todo ello, la Quimiometría se sitúa en un campo interdisciplinario. Muchos autores [2], coinciden en considerarla como una rama aplicada de la Química Analítica y la Estadística aplicada, con una función análoga a la que desempeñan disciplinas como la Biometría,

Sociometría, Econometría o Psicometría en relación a la Biología, Sociología, Economía o Psicología, entre otras. Los fines de la Quimiometría están ligados a la Química y su éxito depende de los problemas químicos que sea capaz de resolver.

La mayoría de las diversas ramas de la Estadística aplicada se desarrollaron en el primer tercio del siglo XX. De esa época datan los primeros números de revistas investigación tales como Biométrica (1901), Econométrica (1933) y Psicométrica (1936). En cambio, el desarrollo de la Quimiometría como disciplina científica independiente no se inició hasta principios de la década del 70. El desarrollo de la Quimiometría fue tardío como consecuencia de la lenta evolución de la instrumentación científica. En la primera mitad del siglo XX era difícil y costoso obtener, tratar y almacenar datos químicos analíticos. Se disponía de pocos datos, y su examen directo permitía al analista extraer toda la información significativa. Esta situación comenzó a cambiar hacia 1960, debido al fuerte impulso que experimentó el desarrollo de nueva instrumentación. Sin embargo, en la década 1960-1970, la recopilación de los datos y su tratamiento manual con algoritmos matemáticos eran procesos lentos y tediosos. Por ello, los algoritmos con cierta complejidad de cálculo (incluida la regresión lineal simple) no tuvieron interés práctico hasta la popularización de las calculadoras electrónicas hacia 1975. Un impulso definitivo de la Quimiometría como disciplina científica independiente fue debido a la popularización del microordenador, y sobre todo su incorporación a los instrumentos científicos, a mediados de la década 1980-1990).

En 1971 el Profesor Svante Wold de la Universidad de Umea Suecia fue el primero que utilizó el término “chemometric” en la argumentación para el financiamiento de un proyecto de investigación [3] y rápidamente su colaboración con el Dr. Bruce Kowalski, de la Universidad de Washington lleva el término a los Estados Unidos.

En la sección de revisión y reseñas de la revista *Analytical Chemistry* de 1972 [4] se reflejaron sólo dos tópicos que muestran de forma incipiente estudios relacionados con la Quimiometría; “El Ajuste de Curvas” y “El Control Estadístico de la Calidad en Procesos Químicos”.

Como parte de los esfuerzos de sus fundadores, para formalizar la Quimiometría, como una disciplina propia dentro de la Química se constituye la Sociedad Internacional de Quimiometría en 1974.

En 1975 se publica el primer artículo que recoge en su título el término “Chemometric”, en el cual su autor el Dr. Kowalski [5] destaca los avances alcanzados; sobre todo, en el área del Reconocimiento de Patrones, de la necesidad de que se reconozca como una nueva disciplina de la Química Analítica, y de la posibilidad real de que revistas de impacto como la *Journal of Chemical Information and Computer Science*, la *Analytical Chemistry* y la *Analytical Chimica Acta* publiquen las investigaciones de esta novel disciplina.

La aceptación de los novedosos procedimientos que se iban abriendo paso en este campo de manera gradual no fue universal - como siempre ocurre-, existieron numerosos escépticos dentro de la comunidad química que planteaban que si se necesitaban complejos procedimientos en el análisis de los datos experimentales esto era una indicación que los experimentos no habían sido diseñados adecuadamente.

No obstante, en 1977 se introduce la sección “Computer Techniques and Optimization” [6] en la revista *Analytica Chimica Acta* que fue la primera revista que publicó artículos dedicados a esta temática. Una vez aceptada la Quimiometría como disciplina independiente la revista *Analytica Chimica Acta* elimina en 1982 la sección “Computer Techniques and Optimization” contando a partir de entonces con su sección propia. [7].

A partir de 1985 surgen dos revistas dedicadas exclusivamente a los aspectos teóricos relacionados con esta temática: Chemometrics and Intelligent Laboratory Systems, y la Journal of Chemometrics.

2 Principales áreas de trabajo

Como expresamos en el epígrafe anterior la Quimiometría se ha desarrollado vertiginosamente a partir de la década de los 90 y consecuentemente su áreas de trabajo se han diversificado de forma incremental para poder adaptarse y dar respuesta a los nuevos retos. Este proceso se ha caracterizado por garantizar dos aspectos básicos: El desarrollo de nuevas teorías y algoritmos para el tratamiento de las datos químicas y el de nuevas aplicaciones de las técnicas quimiométricas en diferentes disciplinas de la Química.

En recopilaciones de la especialidad publicadas en la Analytical Chemistry [8.9.10] se pueden identificar como principales áreas de trabajo las relacionadas con:

- El Reconocimiento de Patrones.
- La Calibración Multivariada.
- La Relación Estructura- Actividad. (QSAR)
- El Procesamiento de Señales Químicas.
- Resolución Matemática de Mezclas Complejas.
- El Diseño de Experimentos.
- El Análisis de Imágenes Químicas.

Como expresamos en la introducción este es el primero de una serie de reportes donde los quimiometristas encontrarán las principales bases teóricas, donde se sustenta cada una de las áreas de trabajo de la especialidad, así como los principales resultados teóricos y aplicados desarrollados en ellas. Corresponde a este número el desarrollo de estos objetivos para el área del Reconocimiento de Patrones en la Quimiometría.

2.1 Reconocimiento de Patrones

Una de las primeras áreas de trabajo que abordó la Quimiometría y donde se recogen importantes resultados es precisamente la relacionada con el Reconocimiento Patrones.

Son innumerables los problemas en Química, que requieren para su solución del empleo de estas herramientas. Por ejemplo ¿Puede ser utilizado un espectro infrarrojo para clasificar compuestos en cetonas y ésteres?, ¿Existe en ese espectro patrones que permitan relacionar la información física con el conocimiento químico?, ¿Puede un cromatograma ser usado en la Química Forense para determinar la causa de un incendio?, ¿Puede ser utilizado un cromatograma para determinar el origen de una droga, y además es posible determinar rasgos en este cromatograma que permita diferenciar drogas de diferentes tipos y orígenes? ¿Es posible utilizar las medidas de concentración de elementos metálicos pesados para determinar la fuente de contaminación en un río? ¿Es posible a través de la imagen de una corrida de muestras de ADN sobre geles de poliacrilamida extraer la información para la conformación de los perfiles correspondientes? Las respuestas a estas interrogantes pueden ser encontradas dentro del ámbito de trabajo de esta importante área de la quimiometría.

Entre los objetivos fundamentales del Reconocimiento de Patrones en la quimiometría tenemos: La identificación de relaciones y/o vínculos entre objetos, en nuestro caso

denominados muestras químicas o grupos de estas previamente caracterizadas a través de diversos métodos de análisis instrumental, que permitan su agrupamiento y clasificación según la similitud de sus espectros, cromatogramas, análisis elemental, imágenes etc.

De igual manera, estas herramientas deben permitir la clasificación de nuevas muestras como pertenecientes a un grupo determinado de una forma rápida y objetiva. [11]

La Quimiometría utiliza principalmente el enfoque estadístico del Reconocimiento de Patrones atendiendo generalmente a la naturaleza numérica de la data, no obstante con el creciente desarrollo de nuevos métodos y equipos de análisis instrumental multivía, que incorporan inclusive técnicas de procesamiento de imágenes y otras que pudiesen generar variables cualitativas, no se descarta en un futuro el empleo de otros enfoques entre ellos el Lógico Combinatorio.

Para el cumplimiento de sus objetivos, el Reconocimiento de Patrones en la Quimiometría emplea diferentes métodos de trabajo los cuales podemos dividir en dos grandes grupos: (1) El análisis exploratorio de datos, (2) Los métodos de clasificación entre los que se encuentran los Supervisados y No Supervisados.

2.1.1 Análisis Exploratorio de Datos

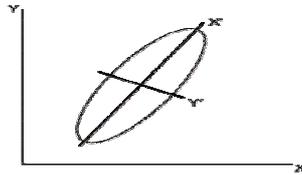
Las técnicas de exploración de datos se utilizan para poner de manifiesto y resaltar la información contenida en una matriz de datos multidimensional.

La observación directa de los datos, así como la utilización de herramientas simples de graficación como son los histogramas, o los gráficos de dispersión sobre pares de variables, son útiles pero insuficientes para poner de manifiesto y describir en toda su magnitud las estructuras presentes.

Técnicas de exploración de datos más sofisticadas y robustas como las incluidas en la rotación propia, seguida del análisis de los resultados obtenidos o “análisis de los componentes principales, se requieren para lograr definir estructuras que ocupan espacios de más de 3 dimensiones, para identificar tendencias ocultas presentes en los datos y estudiar las causas o fuentes de varianza a que obedecen, para lograr obtener las variables latentes y fundamentales del sistema y sobre todo lograr la reducción de dimensiones, de modo que la información relevante contenida en la matriz multidimensional pueda quedar reflejada, del mejor modo posible, sobre dos o tres dimensiones oblicuas obtenidas como combinaciones lineales de las variables originales.

El análisis de componentes principales, es una valiosa herramienta estadística y de proyección, que persigue maximizar la información de la varianza presente en una data y representarla en el menor número de dimensiones posible, utiliza como base la rotación propia de los ejes de la data para conformar nuevos ejes que contienen la mayor información de la varianza posible, esto nos permite buscar de forma secuencial las direcciones del espacio que ofrece la mejor visión posible de la misma, o sea lograr una nueva visualización de la data en el hiperespacio.

Un ejemplo bidimensional se muestra en el siguiente gráfico, donde el nuevo eje de coordenada X' representa la dirección de la máxima varianza o tendencia principal de los datos y representa al primer componente principal mientras el eje Y' es representativo del resto e identifica al 2do componente principal.



Matemáticamente, estas operaciones son consideradas como transformaciones simples del álgebra lineal, donde los nuevos ejes de coordenadas, denominados vectores propios (eigenvectors), auto vectores o componentes principales, los cuales además cumplen la condición de ser ortogonales entre ellos, se construyen como combinaciones lineales de las variables originales. Los componentes principales son variables latentes, esto es, modelan las principales tendencias presentes en la nube de datos.

Otras contribuciones importantes que brinda el análisis de componentes principales, son las relacionadas con la reducción del número de variables necesarias para representar el sistema, proporcionando un nuevo conjunto de variables que describen propiedades encubiertas hasta el momento, de igual manera un análisis de la puntuaciones, cargas y autovalores generados permite conocer cuáles son las variables originales que más aportan a los componentes principales y por consiguiente a las tendencias predominantes de los datos.

De igual valor resulta su aporte en la detección de las variables fundamentales. Esto es factores subyacentes (ambientales, clínicos, industriales, geográficos, históricos etc.) que expliquen la configuración de los diferentes grupos y las correlaciones observadas. [2, 12,13].

2.1.2 Métodos de Clasificación

En un análisis clasificatorio se construyen modelos capaces de pronosticar la pertenencia de un objeto a una categoría sobre la base de las características del objeto.

2.1.2.1 Clasificación No Supervisada

Es aquella clasificación donde no se cuenta con información a priori sobre la existencia de clases entre los objetos. Se tiene entonces como tarea, encontrar las posibles clases existentes en la matriz de datos a partir del agrupamiento de objetos con similares características.

Entre los métodos empleados por la quimiometría, en la clasificación no supervisada se encuentran:

Análisis de Cluster

Es uno de los métodos más utilizados en la clasificación No supervisada y se sustenta o depende de una definición cuantitativa de la similaridad entre pares de objetos o grupos en formación y de lo que representa esa medida para la conformación de los grupos. [11,15]

Todo análisis de cluster consta de varios pasos, el primero, consiste en determinar la similaridad entre objetos, la cual se expresa a través de un indicador numérico. Cuatro de las principales vías que son empleadas para la determinación de esta similaridad son: El cálculo de coeficientes de correlación entre muestras, obtención de la distancia Euclidiana, de la distancia de Manhattan o de la distancia de Mahalanobis.

El siguiente paso está relacionado con la interconexión entre objetos. La técnica más comúnmente utilizada es el llamado cluster por aglomeración donde cada objeto simple es

gradualmente conectado con los otros en grupos. Cualquier de las cuatro medidas de similaridad contempladas en el primer paso puede ser empleada. A excepción de la medida de similitud por coeficiente de correlación, a menor distancia obtenida entre uno y otro objeto mayor es su similaridad. Pongamos un ejemplo para lograr una mayor comprensión. Tenemos una matriz de 6 objetos (1-6) y 5 medidas (A-E) como se muestra en la tabla 1.

Tabla 1. Ejemplo de análisis de Cluster.

	A	B	C	D	E
1	0.9	0.5	0.2	1.6	1.5
2	0.3	0.2	0.6	0.7	0.1
3	0.7	0.2	0.1	0.9	0.1
4	0.1	0.4	1.1	1.3	0.2
5	1.0	0.7	2.0	2.2	0.4
6	0.3	0.1	0.3	0.5	0.1

Utilizamos como medida de similaridad, la determinación de los coeficientes de correlación entre muestras, obteniéndose la siguiente matriz de correlación. (Tabla No.2)

Tabla 2. Matriz de Correlación.

	1	2	3	4	5	6
1	1.000					
2	- 0.041	1.000				
3	0.503	0.490	1.000			
4	- 0.018	0.925	0.257	1.000		
5	- 0.078	0.999	0.452	0.927	1.000	
6	0.264	0.900	0.799	0.724	0.883	1.000

En nuestro caso, se asume que mientras más negativo es el coeficiente de correlación menor similaridad entre las muestras.

A partir de los datos de la tabla 2. debemos encontrar las dos muestras que son más similares, como se puede observar estas son las muestras 2 y 5 ya que tienen entre ellas los mayores coeficientes de correlación (0.999). Con estas dos muestras se forma un grupo o cluster quedando remanentes las muestras 1, 3, 6, 4.

El problema a continuación reside en como representar el nuevo grupo, para esto hay diferentes opciones. La principal tarea ahora consiste en recalcular los valores de similaridad numérica entre el nuevo grupo formado y los objetos remanentes. Existen tres Vías para lograr éste objetivo: (1) Aplicando la técnica de los vecinos más cercanos, (2) aplicando las técnicas de los vecinos más lejanos (3) aplicando el promedio de conexión entre objetos. [11]

A partir del cálculo las nuevas medidas de similaridad, el próximo paso consiste en continuar agrupando las muestras, manteniendo el procedimiento anterior hasta tanto todas estén unidas en

grupos. En la tabla 3. se muestran los resultados de este proceso utilizando como método de conexión de las muestras, el del cálculo de los vecinos más cercanos y como medida de similaridad los coeficientes de correlación.

Tabla No.3 Análisis de Cluster por vecinos más cercanos utilizando coeficientes de correlación como medida de similaridad.

	1	2	3	4	5	6
1	1.000					
2	-0.041	1.000				
3	0.503	0.490	1.000			
4	-0.018	0.925	0.257	1.000		
5	-0.078	0.999	0.452	0.927	1.000	
6	0.264	0.900	0.799	0.724	0.883	1.000

	1	2&5	3	4	6
1	1.000				
2&5	-0.041	1.000			
3	0.503	0.490	1.000		
4	-0.018	0.927	0.257	1.000	
6	0.264	0.900	0.799	0.724	1.000

	1	2&5&4	3	6
1	1.000			
2&5&4	-0.018	1.000		
3	0.503	0.490	1.000	
6	0.264	0.900	0.799	1.000

	1	2&5&4&6	3
1	1.000		
2&5&4&6	0.264	1.000	
3	0.503	0.799	1.000

	1	2&5&4&6&3
1	1.000	
2&5&4&6&3	0.503	1.000

En la tabla se puede observar que los objetos más similares en cada paso, son indicados por negritas y las nuevas medidas de similaridad están sombreadas. En este ejemplo en particular, las muestras 1 y el 3 no tienen una gran similaridad con el grupo principal, en este caso, otros nuevos grupos deben formarse. Es importante entonces además determinar hasta que medida de similaridad una muestra se conecta con el grupo principal y cuales muestras pasarán a conformar otros grupos.

Muchas veces los resultados de los cluster jerárquicos son presentados en forma de **Dendograma**. Los objetos son organizados en filas acorde a su similaridad: El eje vertical representa la medida de similaridad a la cual sucesivamente las muestras son unidas al grupo principal

En la figura 1. se presenta el Dendograma correspondiente al análisis de cluster representado en la tabla 4.20. Se puede observar claramente como el objeto 1 es muy diferente al resto y quedaría por analizar si el objeto No. 3 pertenece o no al cluster inicial atendiendo al límite que se establezca para la conformación del grupo, de acuerdo a su medida de similaridad.

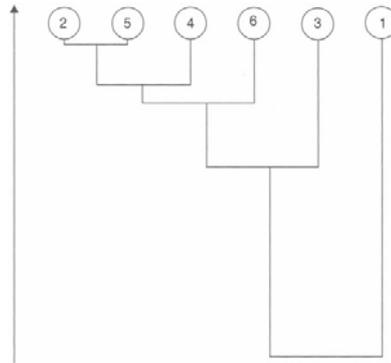


Fig. 1. Dendrograma del análisis del cluster Tabla 2.

Una forma usual de operar consiste en aplicar diferentes métodos de análisis de cluster y comparar sus resultados, si similares cluster son obtenidos podemos estar más seguros de los resultados alcanzados en un la clasificación.

Redes Neuronales Artificiales

Las redes neuronales artificiales son métodos diseñados para el procesado de datos y la organización del conocimiento basado en la imitación del funcionamiento de los sistemas nerviosos biológicos, y es capaz de predecir con precisión diferentes clases. Una Red neuronal no se basa en un modelo algebraico explícito, sino en un conjunto de "unidades" de activación, denominados también "nodos" o "neuronas artificiales" interconectados entre sí en forma de red. Pueden ser utilizadas para resolver problemas de Reconocimiento de Patrones tanto Supervisados como No supervisados.

Como se muestra en la figura.2. las neuronas biológicas tiene cuatro componentes básicos: Dendritas, Soma, Axón, y Sinapsis. A través de las Dendritas, las neuronas reciben señales provenientes de estímulos externos o de otras neuronas, en el Soma las combinas mediante una operación no lineal, y finalmente genera un resultado que se comunica a otras neuronas vecinas a través del Axón y las Sinapsis.

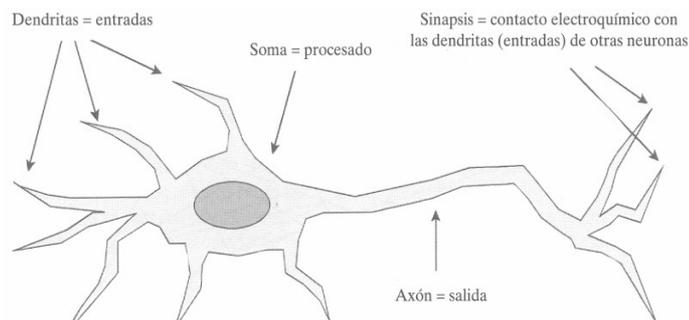


Fig.2. Esquema de una neurona biológica mostrando sus cuatro componentes básicos.

Como se indica en la figura 3. una neurona artificial simula las funciones básicas de la neurona natural. La neurona artificial contiene dos algoritmos, uno de los cuales calcula la suma ponderada de los valores que le llegan por las conexiones de entrada, y el otro denominado "

función de transferencia", genera una respuesta o salida que se comunica a otras neuronas. La red de neuronas es capaz de "aprender", lo que realiza principalmente mediante el ajuste de los "pesos" de las conexiones entre neuronas, hasta que la red en su conjunto proporcionan predicciones con la suficiente precisión.

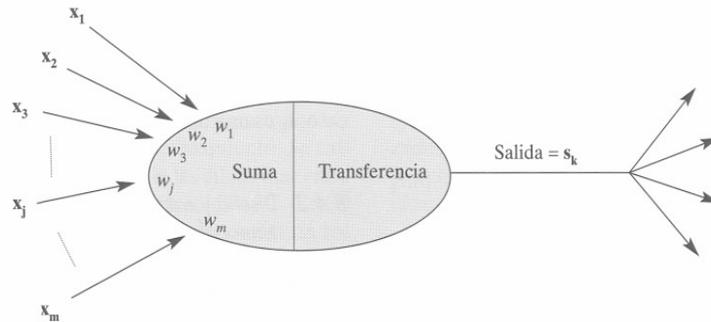


Fig. 3. Esquema de una neurona artificial mostrando sus cuatro funciones básicas: ponderar las entradas, sumarlas, calcular la salida y comunicarla a otras neuronas.

La figura 4. muestra el esquema de una red sencilla, unas neuronas dispuestas en tres filas o capas denominadas: capa de entrada, capa escondida y capa de salida. La capa de entrada tiene tantas entradas como variables manifiestas, y genera a su vez el mismo número de salidas, cada una de las cuales se comunica a todas las neuronas de la capa escondida. La capa escondida suele tener una o dos neuronas más que la capa de entrada. Finalmente, la capa de salida tiene tantas neuronas como clases reconocidas existan en los datos (clasificación supervisada) o como clases que se quieran predecir (clasificación no supervisada).

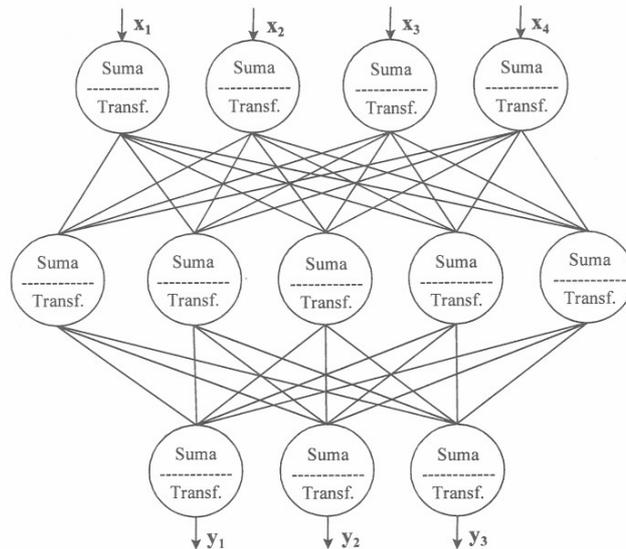


Fig. 4. Esquema de una red neuronal diseñada para predecir tres respuestas a partir de cuatro variables previctorias.

Para que la red funcione correctamente es necesario someterla a una etapa de entrenamiento. Cuando la red está entrenada, las neuronas de la capa de entrada se alimentan con los valores de las variables originales o manifiestas para el nuevo objeto. Los resultados de la capa de entrada se procesan en la capa escondida, y los resultados generados por esta se procesan de nuevo en la capa de salida. En un análisis supervisado los resultados de la capa de salida son las probabilidades de que el nuevo objeto pertenece a cada uno de las categorías

Para los problemas No supervisados en quimiometría, diferentes tipos de redes neuronales pueden emplearse entre las más utilizadas tenemos: a) Las redes neuronales de Kohonen. [16]. b) Las redes neuronales de Teoría de Resonancia Adaptativa (Adaptative resonance theory (ART). [17, 18,19, 20,21].

En las redes neuronales de Kohonen uno de los logros esperados es encontrar grupos de objetos similares. La esencia del algoritmo de Kohonen es la comparación repetida de los vectores que representan los datos de la muestra, con los vectores de peso usando una distancia métrica como por ejemplo la distancia Euclidiana o los coeficientes de correlación. Durante cada comparación simple, un vector ganador entre los vectores de peso es encontrado, el cual tiene el mayor grado de similaridad (menor distancia) con la muestra. El vector de peso ganador es adaptado para estar cerca del vector muestra de entrada.

Junto al ganador, otros vectores de pesos más alejados son modificados alrededor del ganador con una vecindad topologicamente limitada. Al inicio la región es grande pero va disminuyendo lentamente con el tiempo de entrenamiento. Después de la adaptación, el grupo vectores de peso perteneciente al dominio, se hace ligeramente más similar a los vectores de entrada, en términos de la distancia métrica empleada. Una comparación de todos los vectores muestra con todos los vectores de peso y sus modificaciones se denomina un período. Este proceso es repetido sobre muchos periodos lográndose una auto organización de las muestras en una Red neuronal de baja dimensionabilidad. Las muestras, forman una estructura visual relacionada con el espacio original m-dimensional de las variables.

El resultado final de este proceso de entrenamiento es una colección de grupos de vectores muestra agregado con algunos de los vectores de peso. El número de las neuronas logradas es v y puede ser menor o igual al número de neuronas originales. El vector de peso entrenado por este tipo de cluster de agregación a partir de vectores de entrada es numéricamente cercano a su vector principal.

Una de las ventajas que se le asignan este método es que logra representar los resultados a través de una proyección o mapeo que preserva y visualiza más la información. La interpretación de los resultados enriquece con la incorporación de herramientas de visualización como la conocida (The Minimal Spanning Tree [16] que permite la visualización de las relaciones globales entre las neuronas obtenidas en el mapa. .Wienke y Hope [17] brindan una completa descripción de este algoritmo.

Las Redes neuronales ART, fueron introducidas por Grossberg [18,19] como un modelo matemático que fuese capaz de describir importantes funciones del cerebro como, el aprendizaje, el almacenaje de información distribuida y paralela, el reconocimiento de patrones etc.

El principal objetivo de los modelos basados en ART es precisamente entender y describir como el cerebro humano es capaz de identificar un evento inesperado sin tomar en consideración si está presente o no en la base de conocimientos ya existente. Aún más estos modelos tratan de describir la habilidad del cerebro de expandir el conocimiento mediante el aprendizaje de situaciones no conocidas, sin interrumpir o destruir el conocimiento ya almacenado.

El Algoritmo ART-2a [20] es un nuevo tipo de algoritmo desarrollado dentro de los modelos ART que presenta como ventaja que es matemáticamente más simple y computacionalmente más barato que los ART convencionales. La esencia de este algoritmo radica en la formación dinámica de la matriz de pesos, por lo que cada columna de la matriz de peso describe la naturaleza del futuro centroide de la clase. El tamaño de la clase es controlada por un parámetro de vigilancia que se mueve entre 0 y 1.

Este algoritmo es descrito a profundidad [20] y se ha empleado fundamentalmente en investigaciones relacionadas con la quimiometría ambiental sobre todo en la caracterización de nuevas partículas en el aire. [21,22]

Las máquinas de soporte vectorial. (SVM) (Support Vector Machine)

El SVM es una nueva clase de clasificadores, que recién se incorporan en el quehacer de la Quimiometría. Utilizan una variedad de métodos Kernel tales como: Los polinomiales, los de base radial, y los Gaussianos [23,24].

Son métodos para la creación de funciones a partir de un conjunto de datos con entrenamiento previo. Las funciones pueden ser de clasificación y sus salidas son binarias, aunque también pueden generarse funciones de regresión.

Investigaciones recientes en el área del reconocimiento de patrones en la quimiometría utilizando este tipo de clasificadores se reportan por el Grupo de Quimiometría de la Universidad de Bristol UK. a cargo del Dr. Simeone Zomer en temas relacionados con la determinación del origen farmacéutico de medicamentos en tabletas para investigaciones forenses. [25].

2. 1. 2. 2 Clasificación Supervisada

Este tipo de clasificación ha sido utilizado ampliamente a través de los años en Quimiometría. Para construir un modelo de clasificación supervisada es necesario disponer de una muestra de objetos cuya clase sea conocida y para los que también se conozcan los valores de las variables predictoras. La pertenencia de los objetos a las clases puede ser supuesta, esto es, puede tratarse de una hipótesis a comprobar. La asignación de los objetos a las clases debe ser exhaustiva (todos los objetos pertenecen a alguna clase) y mutuamente exclusiva (ningún objeto pertenece a más de una clase). Estos objetos forman lo que se denomina el "conjunto de entrenamiento" con el cual se construye el modelo de clasificación. Una vez construido, el modelo se utiliza para predecir la clase a que pertenecen nuevos objetos a partir de la medida de las variables predictoras. Es usual utilizar muestras adicionales pertenecientes a clases conocidas, que no forman parte del conjunto de entrenamiento original, para comprobar el modelo y controlar su precisión y exactitud. De manera general en una clasificación supervisada se siguen los siguientes pasos: (1) modelación del conjunto de entrenamiento. (2) validación del modelo. (3) mejoramiento de los datos (si es necesario). (4) aplicación del modelo.

Existe una variedad de técnicas de clasificación entre las más conocidas tenemos: El análisis discriminante, el método No paramétrico de los k vecinos más próximos o "método k-NN" (k nearest neighbours), sin embargo otros métodos también son empleados para estos fines entre los que tenemos: El método de la máquina de aprendizaje lineal, los métodos de densidad y potencia, los métodos de modelación de clases (SIMCA y UNEQ) y el sistema de expertos para construcción de reglas ID3. [14]

Análisis Discriminante

En análisis Discriminante se utiliza un algoritmo que busca funciones o vectores discriminantes, esto es, combinaciones lineales (ADL) o cuadráticas (ADC) de las variables originales que maximizan la varianza entre clases, a la vez que minimizan las varianzas inter-clases.

Para construir el modelo es necesario asignar los objetos del conjunto de entrenamiento a una clase dada. Para ello, se añade una variable categórica a la matriz de datos conteniendo tantas clases como sean necesarias. Tomemos un ejemplo: Queremos clasificar el café en grano en dos clases. La variable categórica es "TIPO" que contiene las clases "Robusta" y "Arábica". El (ADL) estima los coeficientes a_1, a_2, \dots, a_m , de la función discriminante lineal $f = a_1 x_1 + a_2 x_2 + \dots + a_m x_m$ que es capaz de predecir la pertenencia de los objetos a una u otra categoría.

Por encima de un cierto valor de f , el objeto queda clasificado como "arábica", y por debajo de dicho valor se considera que la clase es "robusta". Si existen tres clases, es necesario construir dos funciones discriminantes, f_1 y f_2 así, la función f_1 distingue una de las categorías de las otras dos y f_2 distingue las dos categorías que quedan entre sí.

Las funciones Discriminantes se construyen de una en una, buscando las direcciones del espacio que hacen máxima la expresión: $\lambda' = \text{SCD} \setminus \text{SC1}$, donde SCD es la suma de los cuadrados de las distancias Euclídeas entre los objetos que pertenecen a distintas clases, en la dirección que indica la función discriminante y SC1 es la suma de los cuadrados de las distancias Euclídeas entre los objetos que pertenecen a la misma clase, también en la dirección de la función discriminante. A partir de q clases se obtienen $q-1$ funciones discriminantes, en orden decreciente de su valor de λ' , y manteniendo la ortogonalidad entre ellas.

Una vez determinada la función Discriminante es posible establecer fronteras explícitas que separan las regiones del espacio que pertenecen a cada categórica. Para ello, y tal como se indica en la figura 5 cada clase se modela en torno a su centroide mediante una distribución normal de la probabilidad. En ADL, la probabilidad decrece por igual en todas las direcciones del espacio m -dimensional de las variables originales siendo alta para puntos situados cerca del centroide de la clase considerada, y baja para puntos alejados. Las curvas que unen puntos de igual probabilidad son círculos para dos variables originales, esferas para tres, e hiperesferas para 4 o más variables.

La clasificación de objetos se hacen en función de sus probabilidades de pertenencia a una u otra clase: un objeto queda clasificado dentro de la clase para la que se obtiene la mayor probabilidad. La región crítica o frontera que separa dos clases es una recta, o un plano o hiperplano, donde se igualan las probabilidades de pertenencia a una u otra clase, esta línea recta o plano es ortogonal a la función discriminante encontrada, figura 6.

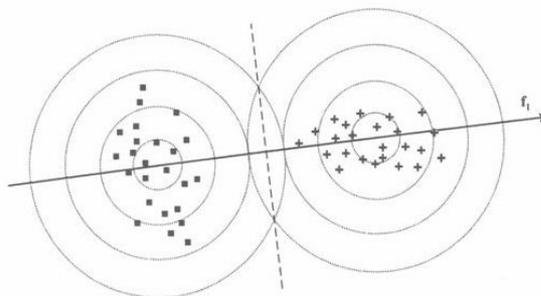


Fig. 5. ADL: curvas circulares de igual probabilidad en torno a los centroides de dos categorías (líneas de puntos), y fronteras entre ambas categorías (línea de trazos).

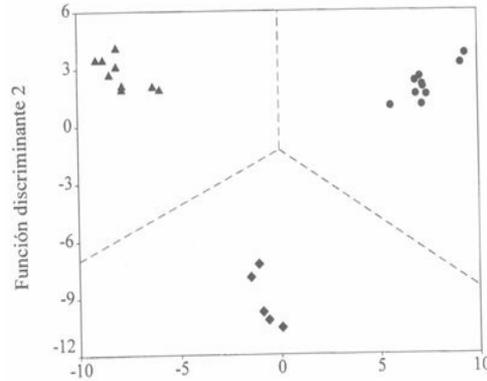


Fig. 6. Gráfico mostrando las fronteras de decisión entre pares de categorías (línea de trazos)

El ADL es adecuado para separar categorías con varianzas similares en todas las direcciones del espacio, sin embargo, cuando las categorías tienen varianzas altas en unas direcciones y bajas en otras, el análisis discriminante cuadrático (ADL) suele dar mejores resultados. También existen casos extremos como los que se muestran en la figura 7, en los que dos clases no son linealmente separables. Cuando una clase se introduce en la ensenada exhibida por otra, cualquier frontera recta o plana clasificará incorrectamente algunos tipos.

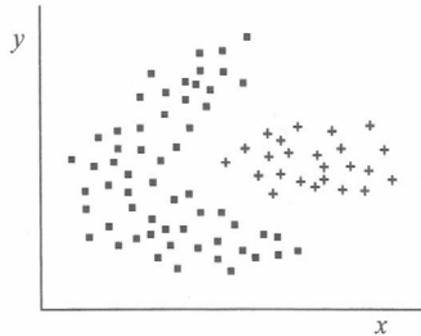


Fig. 7. Categorías no separables linealmente.

Como se indica en la figura 8 en el ADC se utilizan curvas de probabilidad elípticas etc. las curvas elípticas pertenecientes distintas clases se cortan definiendo fronteras curvas.

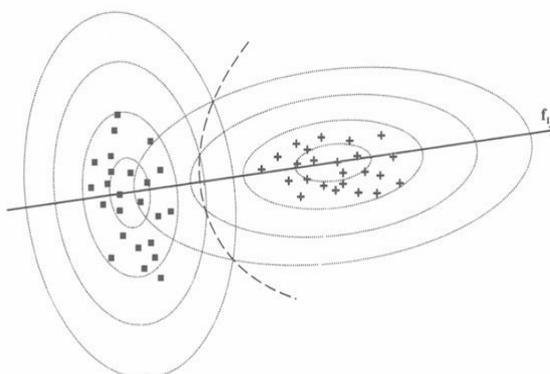


Fig.8. ADC: Curvas elípticas de igual probabilidad en torno a los centroides de dos categorías (líneas de puntos), y frontera entre ambas categorías (línea de trazos).

El Método K-NN

Los métodos no paramétricos de clasificación (no basados en modelos) se basan en la proximidad del objeto nuevo a los objetos del conjunto de entrenamiento. El más utilizado es el de los k - NN, que clasifica al objeto nuevo en la categoría a la que pertenecen sus k vecinos más próximos, siendo k un valor arbitrario, normalmente bajo e impar, 3 o 5, en la figura 9 se esquematiza el funcionamiento de un método 5- NN.

Para aplicar el método se hallan las distancias entre el objeto nuevo y todos los objetos del conjunto de entrenamiento en ocasiones, se utiliza un sistema de puntos ponderados o “votos de distinta calidad”. Una forma frecuente de ponderar los puntos consiste en asignar k puntos a la categoría del vecino más próximo, y así sucesivamente. Se suman los puntos obtenidos por cada clase, y el objeto nuevo se asigna a la clase que obtiene más puntos. En el ejemplo de la tabla No. 4, los vecinos de la categoría A suman 11 puntos, mientras los que pertenecen a la categoría B sólo alcanzan 4 puntos, por tanto el nuevo objeto queda clasificado en la categoría A.

Las distancias pueden calcularse directamente sobre la matriz objetos- variables sin escalar o autoescalada, y también antes o después de realizar una rotación propia, si se realiza la rotación propia se tiene la ventaja de trabajar en el espacio no correlacionado, en donde las distancias Euclídeas tienen significado correcto.

En comparación con el análisis Discriminante el mayor inconveniente del método k -NN es que no calcula las probabilidades de pertenencia del objeto nuevo a cada categoría. Por otra parte, sus ventajas son las siguientes:

a) un solo modelo (no explícito) permite clasificar nuevos objetos independientemente del número de clases. En cambio, en el análisis Discriminante se requiere establecer una frontera (explícita) para cada par de clases.

b) la calidad de las predicciones depende poco de la forma que tengan las clases. Por esta razón, el método puede clasificar correctamente objetos aunque las clases no sean linealmente, ni cuadráticamente separables. El método k -NN establece fronteras irregulares entre las clases.

Máquina de Aprendizaje Lineal (Linear Learnig Machine)

Este método [26] también encuentra fronteras lineales entre las clases. Tales fronteras pueden calcularse a través del sistema de prueba y error en el cual los errores de la prueba inicial permiten que la próxima prueba pueda ser desarrollada más inteligentemente. Para un espacio m - dimensional la superficie de decisión tiene que ser de una dimensión menor que m . Este

método puede ser extendido a la separación de clases múltiples a través de una secuencia de superficies que producen separaciones binarias.

Métodos potenciales o de densidad

En los métodos potenciales [27], se asume que existe un campo potencial alrededor de cada objeto. Tal campo puede ser triangular o Gausiano y está caracterizado por su ancho. La función potencial acumulativa para los objetos se obtiene adicionando las alturas de las funciones individuales de los objetos. Esta función acumulativa es dividida entre el número de objetos para tener la función media y asume un carácter probabilístico. La clasificación de los objetos dentro de una de las clases establecidas puede ser alcanzada determinando la asignación a la clase que produzca el máximo potencial.

Métodos de modelación de clases

En este tipo de métodos, se desarrolla un modelo matemático para cada clase dentro del conjunto de entrenamiento. Tal modelo define un dominio en el espacio y se desarrollan procedimientos para probar si nuevos objetos caen dentro de ese dominio. Si es así, estos objetos son asignados a la clase apropiada. Dentro de este tipo de métodos tenemos:

SIMCA (Soft Independent Modeling of Class Analogy)

SIMCA [28] utiliza un análisis de componentes principales para desarrollar un modelo de cada grupo o clase dentro del conjunto de entrenamiento. Esas clases se modelan por una o series de estructuras lineales (una recta, un plano etc.,) dependiendo del número de componentes requeridos para reproducir los datos de la clase. Es posible definir superficies enlazadas alrededor de estas estructuras lineales basados en los residuos de los datos después de fijados los componentes, nuevos objetos pueden ser cuantitativamente probados por sus miembros en las clases definidas.

UNEQ

En este método, cada grupo es modelado con una distribución normal Multivariante [29]. La distancia de Mahalanobis para cada objeto en una clase dada. La distancia de Mahalanobis, seguida de una distribución de Chi-cuadrada a un intervalo de confianza de un 95% puede ser entonces definida. Este intervalo de confianza representa la frontera para la clase y nuevos objetos pueden ser probados para determinar si caen dentro de la clase definida dentro de esa frontera.

Sistemas de expertos para la construcción de reglas

Estos métodos fueron desarrollados como una de las técnicas utilizadas en la esfera de la inteligencia artificial. La idea consiste en que un conjunto de entrenamiento es presentado a una máquina de inducción basadas en diferentes procedimientos de clasificación y un conjunto de reglas se derivan de estos ejemplos y pueden ser usadas por nuevos objetos. [30]. Uno de los métodos de este tipo más ampliamente utilizados es el algoritmo ID3 [31], que emplea información teórica como base para la separación de clases y ha sido incorporado en diferentes programas comerciales.

Si existe una colección de objetos representativos de diferentes clases, la entropía de la información será alta. La separación de los objetos en grupos homogéneos producirá una baja entropía de la información. Este método separa secuencialmente los objetos utilizando un clasificador univariante que produzca un decremento máximo en la entropía. El proceso continúa

hasta que cada objeto es incluido en un grupo homogéneo. Como resultado de esta clasificación una serie de reglas están disponibles que pueden ser usadas para clasificar objetos del mismo dominio. Por tanto en vez de construir reglas a través de preguntas a expertos, el sistema desarrolla las reglas provenientes de la clasificación del conjunto de entrenamiento.

3 Principales investigaciones llevadas a cabo en el área del reconocimiento de patrones en la quimiometría en el periodo comprendido entre 1995-2004

Como expresamos en el epígrafe. 2 en el proceso de evolución de la Quimiometría se pone de manifiesto un ingente esfuerzo por mantener un balance entre la producción científica teórica y la aplicada, o sea: (1) El desarrollo de nuevas teorías y algoritmos para el tratamiento de las datas químicas (2) El desarrollo de nuevas aplicaciones de las técnicas quimiométricas en diferentes disciplinas de la Química como: la química ambiental, agrícola, forense, médica, y la ingeniería química. No obstante estos esfuerzos, en el periodo que se analiza, los resultados muestran un balance y tendencia favorable en el campo del desarrollo de novedosas aplicaciones que han influido no obstante, indudablemente, a enfrentar nuevos retos en el campo teórico.

Este tema ha sido motivo de múltiples debates y reflexiones en los últimos tiempos, Dr. Svante Wold uno de los padres de la Quimiometría, mantiene de forma permanente en sus consideraciones, el análisis a su pregunta acerca de ¿Cuan exitosa ha sido la Quimiometría? [32]. Wold considera que la Quimiometría tiene éxito en la medida que es capaz de solucionar los problemas químicos que se le encomiendan y no por el número de métodos o de algoritmos que se desarrollen y que los resultados alcanzados hasta el momento han demostrado que la Quimiometría ha sido muy exitosa en la solución de estos problemas y por lo tanto que las preocupaciones existentes en cuanto a la tendencia creciente a las aplicaciones, no es un indicador para considerar un estancamiento en el desarrollo de la Quimiometría pues en última instancia son precisamente estas aplicaciones las que generan en gran parte los retos en el quehacer teórico.

Otros Investigadores prestigiosos de la Quimiometría como los profesores Brown y Lavine de la Universidad de Delaware y Postdam respectivamente, consideran en sus resúmenes sobre las principales tendencias en el campo de la Quimiometría [33, 9] que no es siempre posible centrar sólo el análisis sobre el desarrollo de la especialidad , en la capacidad de ésta, para la solución de los problemas químicos, debido a las limitaciones actuales que presentan un gran número de las técnicas de análisis de datos, utilizadas actualmente por los quimiometristas. Plantean que a medida que el análisis de datos en química se hace más complejo, ejemplo: datos perdidos, errores en la correlación de las medidas, grandes volúmenes de datos, etc. se impone ante los quimiometristas la tarea de modificar estas técnicas o crear nuevas para poder manejar esas nuevas estructuras de datos, y que nos es necesario esperar a que las aplicaciones requieran de estos cambios, si se conocen ya en muchos casos las limitaciones existentes y que esto garantiza el desarrollo de la especialidad y a lograr respuestas a más corto plazo de los problemas que puedan venir.

Otros investigadores como Workman [10] tratan de explicar este comportamiento de los resultados con el hecho que para los químicos, sobre todo los que trabajan en la industria o como profesores, la Quimiometría es considerada demasiado compleja y que tiene más de matemática que de química, tergiversando la verdadera esencia de la especialidad, en este sentido los quimiometristas sobre todos los directamente relacionados con el desarrollo de nuevos métodos

y herramientas y su publicación en revistas de impacto, han tenido que romper muchas barreras para que esta especialidad sea apreciada y aplicada en toda su dimensión y por tal motivo se han volcado a lograr que el resultado de sus trabajos se vean reflejados en la aplicación y solución de problemas concretos en todas las esferas del accionar de la química, dando como resultado un número creciente de publicaciones en el campo de las aplicaciones.

Según nuestro punto de vista, ambos factores: las aplicaciones, así como el creación de nuevos métodos son fuentes de desarrollo de la especialidad; ambos son ingredientes indispensables que hay que tener en cuenta para cualquier análisis que se quiera realizar sobre su desarrollo futuro, y que éste como le mostraremos a continuación está garantizado en el presente siglo XXI, ya que la Quimiometría se sitúa en una posición privilegiada para poder resolver una gran variedad de importantes retos en temas relacionados con el análisis de imágenes químicas, el desarrollo del procesamiento de señales de sensores, el aprendizaje de máquinas, el procesamiento multivariado (química combinatoria), la quimiometría-informática (procesamiento de grandes bases de datos químicos) y la exploración de datos multivariados en matrices de datos de alta dimensionalidad, todos estrechamente vinculados y dentro del radio de acción de las herramientas del Reconocimiento de Patrones y la Minería de Datos.

Siendo consecuente con nuestro punto de vista este reporte abordará el tema de las principales investigaciones en el campo del Reconocimiento de Patrones en la Quimiometría en el período seleccionado, tomando en consideración los resultados tanto de las investigaciones fundamentales o teóricas como en las aplicadas.

La Dirección fundamental de las investigaciones en el contexto teórico y práctico sobre Reconocimiento de Patrones, ha estado dirigida hacia el desarrollo de métodos cada vez más sofisticados como resultado del aumento de la complejidad de los problemas en el análisis de los datos. Muchos de estos problemas involucran la modelación de sistemas químicos no lineales.

Una de las líneas de trabajo con mayores resultados en el período es la del análisis exploratorio de datos fundamentalmente relacionados con: La selección de rasgos y las modificaciones y adaptaciones en el análisis de componentes principales aplicados a matrices de alta dimensión, así como el procesamiento de señales espectroscópicas y cromatográficas, interés especial se le presta al desarrollo de métodos para el procesamiento de datos obtenidos por sistemas instrumentales combinados o multivariados. A partir de que los problemas del análisis de datos se hacen más complejos, se necesita que los quimiometristas modifiquen o creen nuevos métodos para poder manejar estas nuevas estructuras de datos.

Entre las principales modificaciones y nuevos desarrollos en el análisis de los componentes principales tenemos: Wentzell [34] ha desarrollado una rutina para el análisis de componentes principales que utiliza la teoría de máxima verosimilitud para compensar los errores de correlación de las medidas entre las variables en la matriz de datos. Manne [35] desarrolló un algoritmo de descomposición de valores para el cómputo de valores perdidos cuando se realiza el análisis de componentes principales en un conjunto de datos ruidosos. Paterro y Hope [36] obtuvieron una nueva rutina para el análisis de componentes principales que logra pesar de forma individual los puntos de los datos.

Los dos resultados de mayor impacto en esta temática estuvieron dirigidos al tratamiento y exploración de grandes matrices de datos donde residen las principales dificultades y problemas no resueltos, sobre todo al enfrentar los datos provenientes de los sistemas bioquímicos (Omic-Proteomics) estos son:

La Transformada de Componentes Principales (Principal Components Transform. PCT), nueva herramienta para el tratamiento de matrices de datos de gran dimensión. [37].

La Transformada de Componentes Principales Segmentada (seg. PCT) nueva herramienta para la exploración de grandes matrices de datos a partir de su tratamiento por segmentos. (matriz segmentada, menor dimensión).[38]

La Transformada de Componentes Principales es una herramienta conceptualmente similar a la Transformada de Fourier, pero que además incorpora la descomposición completa de la matriz a partir de la utilización del Algoritmo NIPALS, seguido del empleo de otros métodos de análisis multivariante convencionales.

El aporte de esta herramienta radica en su capacidad para acelerar y simplificar la aplicación de modelos multivariados a conjunto de datos muy grandes dentro de un razonable tiempo computacional, sin modificar las propiedades estadísticas de estos métodos. Entre las propiedades interesantes de la PCT se encuentra su invarianza ante la variabilidad total de cada muestra cuando se pasa de un dominio (el de las variables originales) hacia el nuevo dominio (espacio de componentes principales) lo que permite usar la validación cruzada en este último. De forma similar la variabilidad total en un subconjunto de las variables originales es conservada después de la aplicación de la PCT lo que significa que ésta puede ser aplicada sin contratiempos a subconjuntos o segmentos de la matriz sobre todo aquellos que contienen información valiosa.

Esta propiedad dio origen al segundo aporte el de la Transformada de Componentes Principales Segmentada. (Seg PCT-PCA). Este logro permite aplicar el Análisis de Componentes Principales (ACP) en segmentos de la matriz de alta dimensión facilitando su procesamiento e interpretación ya que esta herramienta utiliza mucha menos memoria que cuando se aplica a toda la matriz original, ya que la descomposición es aplicada a matrices más pequeñas con la consecuente reducción de los requerimientos de memoria.

La Seg PCT-PCA brinda resultados similares a si se emplea el ACP a la matriz completa utilizando menos recursos computacionales y con mucha mayor rapidez y menos complicaciones tecnológicas. Un ejemplo de esto es el tratamiento de matrices de datos de Resonancia Magnética Nuclear de alta dimensionalidad donde se ha visto que es casi imposible un ACP global y donde se ha podido lograr su aplicación a través de la utilización de la Seg PCT-PCA. Otra ventaja de esta nueva herramienta es que no es necesaria la lectura completa de la matriz inicial en la memoria y desde el momento que los cálculos intermedios de las matrices segmentadas son independientes abre un campo de aplicaciones importantes para la programación paralela.

La temática sobre la selección de rasgos ha sido retomada con nuevos bríos en este período, enfocándose las principales investigaciones hacia la selección de rasgos o descriptores en estructuras químicas que permitan relacionarlas con propiedades químicas o físicas extraíbles o medibles de sus correspondientes espectros o cromatogramas. La mayoría de los estudios se ha centrado en lograr la representación de la estructura de un compuesto a través de una serie de descriptores moleculares y aplicación de un modelo suave que permita descubrir relaciones estructura- propiedad y posteriormente pasar a opciones de clasificación de dichos compuestos.[9].

El desarrollo y la aplicación de las redes neuronales y los cluster han sido las técnicas más empleadas para la solución de los problemas planteados en esta línea de trabajo, con un campo grande de aplicaciones en la química forense y farmacéutica.

Harrington [39] diseñó una red neuronal para predecir la toxicidad de pesticidas órgano fosforado logrando previamente identificar como rasgo fundamental la subestructura activa en

este tipo de compuesto. Heravi [40] utilizó una red neuronal para modelar los factores de respuesta del proceso de ionización por llama de una variedad de compuestos orgánicos. Barden [41] empleo por primera vez una red neuronal para predecir la localización de un pico IR característico para un tipo de compuesto a partir del análisis de 1116 espectros IR de diferentes compuestos orgánicos. Liu [42], Diseñó y entrenó una nueva red neuronal para predecir la constante de asociación de compuestos aromáticos mono y di sustituidos, muy en específico para el caso de la inclusión en el anillo aromático de complejos derivados de la α ciclodextrina. Los descriptores seleccionados y la aplicación del modelo lograron descubrir relaciones valiosas entre las estructuras y las características hidrofóbicas y el tamaño de los sustituyentes en el anillo aromático.

Tetteh [43] entrenó una red neuronal radial para modelar de forma simultánea el punto de flameo y ebullición de compuestos orgánicos utilizando descriptores topológicos. Hall [44] por otra parte, desarrolló una red neuronal utilizando descriptores topológicos para lo cual previamente confeccionó un índice del estado electro topológico de compuestos seleccionados para modelar los puntos de ebullición de un conjunto de 372 compuestos saturados como alcanos, alcoholes, y cloro alcanos. Un método para la predicción de la solubilidad acuosa de 211 drogas fue desarrollado utilizando descriptores topológicos y una red neuronal artificial como modelo [45].

En la utilización de modelos basados en cluster, Wold [46] desarrolló una estrategia para el empleo del análisis de cluster en compuestos orgánicos utilizando datos obtenidos a través de instrumentos combinados y aplicó el mismo con excelentes resultados en la clasificación de 627 alcoholes, previamente cada compuesto había sido caracterizado por 50 descriptores moleculares. La aplicación posterior de un modelo de mínimos cuadrados parciales confirmó la relevancia química del cluster utilizado.

Barros [47, 48] desarrollo un cluster de mínimos cuadrados parciales para la selección de variables. Esta técnica nos brinda una mejor comprensión de diversos diagnósticos obtenidos aplicando clusters a partir del estudio de los dendogramas y del examen de varias propiedades estadísticas asociados con los nodos. En cada nodo es posible recuperar todos los vectores de regresión común para los mínimos cuadrados parciales. Tomando las ventajas de esta técnica y de los rasgos o características que pueden ser extraídos del dendograma, es posible aplicarla para la selección de variables en orden de mejorar las separaciones del cluster y para una mejor comprensión del mismo.

De importancia fueron los resultados alcanzados por Lanteri y Armanino [51] en la obtención y aplicación de dos nuevos métodos IPW-PLS (Iterative Predictors Weighting Partial Least Squares) y UVE (Uninformative Variables Eliminación) para la selección de variables en espectros del infrarrojo cercano de semillas de Brassica Juncea para su posterior clasificación.

Gasteiger [49], utilizó un análisis de componentes principales y una red neuronal SON para caracterizar los cambios electrónicos en el átomo de oxígeno en el escenario de una reacción, a partir de la selección de descriptores característicos. Buenas correlaciones se encontraron entre las similitudes de los cambios electrónicos obtenidos a partir de los descriptores seleccionados y las similitudes en las transformaciones ocurridas a nivel subestructural del átomo en el momento de la reacción. Meusinger [50] construyó un conjunto de reglas de asociación estructura-propiedad para modelar el comportamiento degradativo de 241 compuestos orgánicos utilizando descriptores topológicos y algoritmos genéticos no binarios.

Por último recientemente Varmuza [52,53] logró la modelación de los índices de retención de tres grupos de compuestos de interés toxicológico, a partir de la determinación de un alto número de descriptores moleculares obtenidos con la ayuda del Software Dragón [54].

Atención priorizada a recibido las investigaciones relacionadas con la temática **del procesamiento de señales espectroscópicas y cromatográficas**, Esta línea de trabajo esta indisolublemente vinculada con el proceso de Reconocimiento de Patrones en la Quimiometría, pues muchos la consideran con derecho como una antesala imprescindible de preparación y mejoramiento de la mediciones analíticas provenientes de diferentes equipos instrumentales para obtener información química o física relevante y de calidad que garantice un mejor procesos de clasificación.[8]

Las principales investigaciones han estado dirigidas hacia el desarrollo y aplicación de técnicas tales como los paquetes “wavelet” y las de máxima entropía para la deconvolución de los picos de las señales espectroscópicas y cromatográficas, así como la obtención de métodos de filtrado digital para el mejoramiento y optimización de las señales y las correcciones del fondo y la línea base.

De verdadera explosión puede catalogarse el número de referencias recogidas en la base de datos de las Chemical Abstracts sobre el empleo de las “wavelet” que va desde aplicaciones para el preprocesamiento de espectros infrarrojos [55], la eliminación de ruidos y la compresión de señales por sus bordes [56, 57,58], en apoyo al Reconocimiento de Patrones y la compresión de datos [59] y para el análisis cuantitativo basado en modelos lineales [60, 61,73]

Otros resultados de interés en este campo son los consignados por: Bakshi [62] en la compresión de datos y la extracción de rasgos con wavelets. Kimble [63] Presentó una comparación entre las wavelets y la transformada rápida de Fourier para el análisis de datos dinámicos. La eliminación de interferencias en picos Gaussianos fue estudiado por Mittermayr [57], mientras que Wolkenstein [64], compara el filtrado con wavelets, con técnicas establecidas para la eliminación de ruidos en imágenes EPMA, consignando sus ventajas. Kosanovich [65] examinó una nueva familia de wavelets derivada de la distribución de Poisson para el análisis de señales dependientes del tiempo. Massart [61,59] empleó wavelets para extraer componentes relevantes de espectros infrarrojos con fines de Reconocimiento de Patrones. Liu [66] comprimió espectros IR utilizando redes neuronales con wavelets y Neue [67] Comprimió espectros de resonancia magnética nuclear utilizando la transformada de wavelet. Tanto la transformada rápida de wavelet [68] como la transformada de wavelet B-spline [69] han sido utilizadas para la compresión de espectros UV/VIS. Las wavelets han sido también empleadas en el análisis de señales de espectros de masa para la detección de eventos ocultos. [70]. Interesantes resultan los resultados en la aplicación de la transformada de wavelet en espectros de emisión acústica [71,72] y en el análisis de secuencias de ADN. [74]

Aparecen también consignados en el período los reportes referidos a las técnicas de máxima entropía. Cooke [75] y Dowset [76] aplicaron la deconvolución de máxima entropía a datos de espectroscopia de masa y Zhang la utilizó para el mejoramiento de la resolución de espectros de masa de biomoléculas [77]. Schmieder reportó ventajas en la aplicación de las técnicas de máxima entropía en la reconstrucción de espectros [78].

Con respecto al desarrollo y empleo de técnicas de filtrado digital, Cross [79] brinda una visión de la aplicación de métodos de filtraje para el mejoramiento de espectros de Resonancia Magnética Nuclear. Riris [80] compara el rendimiento de la aplicación de los filtros de Wiener, Kalman y de macheo en el mejoramiento de la sensibilidad y precisión de espectros IR.

Como enunciamos en el epígrafe anterior, una de los principales objetivos del Reconocimiento de Patrones es precisamente el **proceso de clasificación**. En el período objeto a análisis se han reportado el desarrollo de nuevos métodos de clasificación destacándose el uso de las redes neuronales.

Jiang [81] y McAvoy [82] publicaron el desarrollo de un nuevo algoritmo de redes neuronales para el análisis de componentes principales no lineales, el logro integra las curvas principales y la red neuronal para obtener un algoritmo que produce scores y loadings. Chen [83] desarrollo un procedimiento denominado análisis correlativo de componentes para identificar las características de clasificación de la data, con un comportamiento similar a un clasificador de patrones que utiliza una red neuronal. Sagrario [84] desarrolló un algoritmo genético para entrenar una red neuronal que permitiera clasificar la data base, garantizando una optimización directa del número de clasificaciones correctas e identificando claramente las fallidas. Cleij [85] desarrolló un método de proyección lineal basado en una red neuronal de paso atrás (feed-forward) para obtener un mapeo de los datos que maximiza la separación entre las clases predefinidas en un conjunto de datos.

Salenicks [86] empleó un análisis de vector politópico para desarrollar un conjunto base no ortogonal que puede ser utilizado como variables vectoriales para el proceso de Reconocimiento de Patrones en las llamadas datos dinámicas. Jiang [87] por otra parte extendió la extracción de variables a conjuntos de datos no lineales utilizando una red neuronal. El empleo del filtro Kalman resultó novedoso como un instrumento para la extracción de rasgos y como clasificador para espectros en el infrarrojo cercano. [88]

Más recientemente Li [89] desarrolló rutina robusta para el análisis discriminante lineal, la cual aporta grandes ventajas en el tratamiento de "outliers". Carrieri [90] desarrolló una red neuronal para la detección de amino ácidos, azúcares, y otros compuestos orgánicos sólidos a través del empleo de los valores propios de su comportamiento en la difracción de la luz polarizada representados en una matriz de Mueller. El resultado del entrenamiento es una matriz de pesos, que cuando se aplica como filtro, es capaz de discernir entre diferentes tipo de analitos, provenientes de sus compuestos iniciales representados en la matriz de Muller. Lavine [91,92] desarrolló un algoritmo genético para ser empleado en el Reconocimiento de Patrones de datos espectroscópicas y cromatográficas. El algoritmo genético obtenido permite la selección de variables que optimiza la separación de las clases cuando se plotean los dos componentes principales de la data. Debido a que los dos componentes principales capturan el grueso de la varianza en la data, las variables seleccionadas por el algoritmo genético contienen información básica sobre las diferencias entre las clases en el conjunto de datos. Por eso, la rutina para el análisis de los componentes principales embebida en la función fijada por el algoritmo genético, actúa como un filtro de la información, principalmente en la reducción del tamaño del espacio de búsqueda, ya que restringe la búsqueda a un subconjunto de variables donde el ploteo de sus componentes principales muestran una separación sobre la base de clases. Además el algoritmo se concentra en aquellas clases y muestras que son difíciles de clasificar. Las muestras que son consistentemente clasificadas correctamente no reciben un peso elevado a la hora del análisis, en comparación con aquellas que son más difíciles de clasificar. Con el tiempo el algoritmo aprende y alcanza su máxima eficiencia, en una manera similar a como lo realiza una red neuronal. El algoritmo propuesto integra aspectos relacionados con la inteligencia artificial y de la computación evolutiva para brindar un inteligente procedimiento de un paso para el Reconocimiento de Patrones.

Recientemente Marini [93] ha presentado el uso de las redes de neuronales de Kohonen como una herramienta para el modelado de clases. El modelado de clase difiere de las técnicas de clasificación puras, en que se concentran en una clase simple a un tiempo dado y trata de definir el criterio de acceso de cuando una muestra pertenece o no a esa clase en específico. A pesar del amplio uso de las redes neuronales en el Reconocimiento de Patrones estas no han sido utilizadas como tal, para el modelado de clase. En particular las redes neuronales de Kohonen han sido generalmente utilizadas como una herramienta de clasificación no supervisada con posibilidades de mapeo. En este reporte Marini presenta una nueva arista para la utilización de las redes neuronales de Kohonen como herramientas para el modelado de clase. En comparación con el bien conocido método SIMCA para el modelado de clase, esta novel herramienta, reporta ventajas en la visualización de los resultados al explotar las ventajas de mapeado no lineal en 2D típico de las redes neuronales de KOHONEN.

Otro concepto interesante abordado en el periodo esta relacionado con la combinación de clasificadores, en este sentido Zhang y Massart [94] presentan la clasificación de drogas Anti HIV utilizando "Boosting methods" en este caso específico el AdaBoost (Adaptative Boosting) y el método CART (Classification and regresión trees). En AdaBoost varios árboles CART son secuencialmente generados para la clasificación de un conjunto de muestras obtenidas de un conjunto de entrenamiento a través de un método de re muestreo. Las muestras que son fácilmente bien clasificadas le son asignados pesos mayores. Muestras con valores altos de peso tiene mayor probabilidad de ser seleccionadas en el siguiente re muestreo para generar el próximo árbol CART.

La regla de decisión eventual se basa en el voto mayoritario de pesos de todos los árboles. El resultado presentado demostró que AdaBoost mejora la clasificación alcanzada por CART.

Recientemente, en el 2004 han comenzado los reportes relacionados con el empleo del método SVMs (Support Vector Machines) como clasificador en quimiometría, Zomer [95] introduce el tema, brindándonos una visión de sus potencialidades. Plantea que la utilización de este clasificador está dirigido principalmente a la solución de tareas que presentan dificultades, particularmente cuando los límites entre las clases se tornan complejos y el conjunto de datos está caracterizado por un número grande de variables. Entre las principales ventajas que le atribuye se encuentran: a) Su estructura es automáticamente determinada sobre la base del conjunto de entrenamiento y relativamente pocos parámetros se necesitan para su instrumentación. b) El entrenamiento comprende la optimización de una función relacionada con un problema computacional cuadrático convexo, generando por lo tanto una solución completamente reproducible (una de las mayores desventajas que se presentan cuando se utilizan las redes neuronales). c) SVMs es robusto cuando acciona con datos de alta dimensionalidad y el trabajo computacional se logra hacer eficiente a través de la utilización de artificios con estructuras KERNEL. d) su basamento teórico se justifica a partir de la teoría estadística de aprendizaje y el sobreajuste (overfitting) puede ser evitado sin tener que utilizar un conjunto de validación.

No obstante algunas desventajas le son atribuidas: a) Esta solamente concebido para los problemas entre dos clases. b) No se ha utilizado como un método de modelación de clase, que significa que cualquier objeto, aún un claro valor atípico (outlier) puede ser asignado a una clase.

Finalmente, se mantiene aún como una pregunta abierta, ¿Cómo seleccionar el Kernel más adecuado?

En la solución de estas desventajas se encuentran actualmente los principales esfuerzos investigativos con respecto a esta temática.

Los métodos de Reconocimiento de Patrones se han convertido en parte integral de los sistemas artificiales de detección del olor y el sabor comúnmente conocidos como narices y lenguas electrónicas. Estos sistemas en su concepción se conforman a partir del desarrollo de 3 líneas de trabajo: La detección y conformación de los complejos de sensores para la captación de los patrones de señales olorosas o del gusto, el desarrollo y aplicación de las técnicas de Reconocimiento de Patrones para el procesamiento y clasificación de las señales obtenidas y lograr extraer con eficiencia la información requerida y el desarrollo del hardware adecuado para la conformación de los equipos que contendrán e integrarán las diferentes partes del sistema. En este reporte nos centraremos en el análisis del punto b por ser la temática objeto de nuestro estudio.

Ryman [96] reportó las mejoras en el reconocimiento del olor a través de complejo de sensores, utilizando redes neuronales para el procesamiento de las señales. Wang [97] combinó una red neuronal de base radial con el algoritmo “fussy c-means” para mejorar la clasificación de patrones de olor con el empleo de narices electrónicas. La utilización combinada de redes neuronales y la lógica Fussy ha demostrado ser más eficiente en el proceso de reconocimiento del olor que las técnicas convencionales, usando narices electrónicas [98, 99]. Una red neuronal híbrida para la clasificación de olores que aprovecha los beneficios de los mapas auto organizativo de variables fue desarrollada por Di.Natale con resultados satisfactorios. [100].

Por otra parte Menzel [101] desarrolló un modelo dinámico de clasificación de gases para narices electrónicas que alcanza una discriminación eficiente. El modelo combina la clasificación de estados estacionarios con la de análisis en serie de carácter transitorio.

La utilización de las técnicas de Reconocimiento de Patrones se ha puesto también de manifiesto en la selección o comprobación del funcionamiento del complejo de sensores. Porter [102] comparó la eficiencia de un algoritmo basado en árboles de decisión Versus el algoritmo conocido como de escalamiento de montaña para la selección de los componentes de un complejo de sensores destinado para la identificación de componentes volátiles orgánicos. AKybaryan [103] desarrolló un método para la identificación del mal funcionamiento de sensores. Los patrones de entrada pertenecientes a una clase son primeramente escalados para unir la varianza y el punto, o de la media, un proceso de filtrado digital es aplicado para eliminar los segmentos ruidosos, seguido de la aplicación de un método basado en wavelet para cada sensor con el objetivo de extraer las variables de patrones. Un análisis de componentes principales es entonces aplicado para la caracterización de la información contenida en las variables patrones. Doynaz [104] abordó el mismo problema utilizando la estadística de Hotellingt y análisis de componentes principales con resultados también satisfactorios.

Recientemente Skov, [105] desarrolló un nuevo concepto para el manejo de datos provenientes de narices electrónicas. Generalmente los datos generados por las narices electrónicas tienen la posibilidad de ser arreglados en más de 2 dimensiones (muestra x tiempo x sensores), pero sin embargo en su procesamiento y específicamente en el caso de la extracción de variables, generalmente se simplifica el ordenamiento de los datos en una matriz de 2 dimensiones, dado a las limitaciones que presentan muchos software comerciales integrados en los equipos narices electrónicas. El aporte de SKOV en su trabajo consistió en investigar la posibilidad de la utilización de los modelos de procesamiento de datos multivía como los del modelo PARAFAC-2 que permite modelar los datos de narices electrónicas en una manera más adecuada que el convencional tratamiento con un arreglo de 2 dimensiones.

La aplicación de este tipo de modelo demostró ser capaz de brindar una nueva e interesante parametrización de datos provenientes de narices electrónicas, lográndose rescatar una mayor

información de significación para el proceso de clasificación que se perdía con la utilización de los arreglos bidimensionales.

Un análisis de esta temática nos induce a plantear que las nuevas investigaciones en esta temática estarán dirigidas a mejorar la eficiencia del proceso de clasificación de los patrones provenientes de las narices y lenguas electrónicas, a través del desarrollo del concepto. La combinación de clasificadores, sobre todo incorporando la lógica Fussy, así como el tratar de materializar estos modelos de clasificación utilizando los arreglos multidimensionales.

Una de los más importantes campos desarrollados del Reconocimiento de Patrones es precisamente en el campo del análisis de imágenes, sin embargo en la rama química esta línea de trabajo alcanza su real dimensión a partir de finales del 98 con el acople de los equipos de espectroscopia molecular con sistemas de captaciones y procesamiento de imágenes digitales. [10].

Penetrar en el análisis de la estructura Química de las sustancias compuestas y sus mezclas a través no sólo de su análisis espectral sino también con el apoyo del análisis de imágenes y lograr establecer su relación y extraer información valiosa constituye uno de los retos del Reconocimiento de Patrones en esta esfera.

El logro incluye el desarrollo de herramientas, este objetivo para la interpretación, creación o extracción de imágenes virtuales de la data real, la comprensión y posible recuperación de datos, el mejoramiento de las imágenes y la visualización y procesamiento 3D de las imágenes de las estructuras y sus mezclas.

Los conjuntos de datos generados por imágenes Químicas son grandes y multivariados y requieren de un procesamiento significativo. Shweitzer [106] presentó un estudio sobre las limitaciones de las técnicas quimiométricas convencionales para procesar y visualizar los datos de las imágenes. Las tareas de segmentación y clasificación resultaron bastante insatisfactorias en su empleo determinado por la alta dimensionabilidad de la data.

Wold (107) investigó 2 métodos de extracción de variables para la clasificación de imágenes con respecto a tipos de proteínas. Geladi (108) comparó la aplicación de las mínimas cuadrados parciales en imágenes con respecto a las aplicaciones de las técnicas de análisis de componentes principales. Descubrió que las imágenes tratadas por mínimos cuadrados parciales sufre una descomposición diferente a cuando es tratada por componentes principales y que estas especificaciones deben tenerse muy en cuenta.

Buydens, ha realizado aportes novedosos en la aplicación de técnicas de Reconocimiento de Patrones en imágenes provenientes de la Resonancia Magnética (Magnetic Resonance Imaging MRI) y la Resonancia Magnética Espectroscópica y Protónica (Proton Magnetic Resonance Spectroscopic Imaging 1H-MRSI) para la clasificación de tejidos cancerosos. Los espectros registrados contienen importante información Química sobre la especificidad de los metabolitos presentes, información que combinada con el análisis de imágenes, permite extraer información clasificatoria de mayor relevancia [109].

Otras investigaciones interesantes incluyen el uso de la auto fluorescencia espectroscópica y la microscopía de fuerza atómica (AFM) para la clasificación de tejidos humanos, pelos, polvos y diferentes alimentos, combinando los datos provenientes de los espectros de auto fluorescencia con las imágenes aplicando la misma técnica [110, 111, 112, 113].

Por otra parte, Esbensen [114, 115] evaluó la complejidad de las señales como función de una escala geométrica de posiciones locales o globales en imágenes espectroscópicas, utilizando la llamada técnica de medida de ángulo (AMT) para transformar los datos de la imagen. El análisis

de componentes principales de espectros AMT dio dos de los datos de la imagen, fue muy útil en la discriminación de diferentes tipos de polvos.

El estudio y comparación de la aplicación de diversas técnicas de filtraje, permitió a Kazmanec y Walzack [116] lograr encontrar las vías idóneas para mejorar las imágenes de corridas de proteínas sobre geles de poliacrilamida para la posterior clasificación.

El análisis multivariado de imágenes químicas (MIA) irrumpe en el 2004 con grandes retos en la solución de problemas relacionados con la espectroscopia Raman y la espectroscopia en el infrarrojo cercano. La utilización del método de (Evolving Window Factor Análisis, EWFA) brinda grandes resultados en la resolución de curvas espectroscópicas. (117).

Enumerar las investigaciones aplicadas realizadas en el período sería interminable, en tal sentido en el anexo 1 se brinda un compendio bibliográfico de las principales publicaciones registradas en el período, y que puede servir como material de consulta al lector.

A modo de conclusión de este epígrafe, podemos señalar que el área de Reconocimiento de Patrones en la Quimiometría está en una posición de franco desarrollo y que tiene ante sí el reto de dar solución a un conjunto de problemas tanto de carácter teórico como práctico en el presente siglo XXI.

Las principales investigaciones estarán dirigidas a los problemas más complejos y aún por resolver, así como al desarrollo de aplicaciones novedosas.

En el transcurso de la presente investigación hemos podido detectar alguno de ellos, entre los que se encuentran:

- Análisis exploratorio de datos en matrices de alta dimensión (omics y proteomics). Desarrollo de nuevos métodos basados en el principio de la segmentación de la matriz y perfeccionamiento de las ya existentes. (PCT, Seg PCT – PCA)
- La representación de las estructuras químicas de los compuestos a través de descriptores moleculares sigue siendo un problema abierto y complejo, sobre todo para el caso de moléculas grandes como los omics y proteomics (Arreglos de proteínas de ADN). Nuevos tipos de descriptores y su selección deben continuar desarrollándose en los próximos años sobre todo de los tipos BCUT y GRID que son fácilmente comprensibles y computables
- El desarrollo de nuevos algoritmos para relacionar estos descriptores moleculares complejos con el comportamiento de la molécula, lo que es denominado estructura-propiedad-actividad, sigue siendo un problema de intenso estudio de hecho cada día surgen nuevos descriptores para el ADN y la necesidad de relacionarlo con el comportamiento genético futuro de los seres humanos.
- La separación de mezclas químicas a través del análisis de sus espectros, la deconvolución de los picos espectroscópicos y cromatográficos sobre todo en sistemas combinados (GC-MC, DAT-HPLC, etc.) constituye tema candente de trabajo. El desarrollo de nuevos filtros o sus modificaciones, así como la obtención de nuevas familias o modificaciones en la esfera de los wavelet sobre todo en el proceso de extracción de rasgos y en el análisis clasificatorio continúan siendo de interés.
- En el campo de los clasificadores, se enfocan las investigaciones hacia el mejoramiento de la eficiencia de la clasificación con el empleo de nuevos métodos bajo la concepción de la combinación de clasificadores, los Boosting Methods y la combinación de redes neuronales con algoritmos Fussy parecen aumentar categóricamente la eficiencia en el campo de las narices electrónicas.

- Nuevos clasificadores en el campo de la Quimiometría tiene un amplio campo de acción entre ellos los Support Vector Machines, muestran ya sus potencialidades, la asimilación de esta herramienta y su aplicación resulta interesante así como las investigaciones tendientes a suplir sus desventajas como la no detección de outlier, estos limitados a 2 clases y no conocerse aún sus posibilidades de ser utilizado para el modelado de clase.
- La profundización del nuevo concepto introducido por SKOV sobre la utilización de los modelos de procesamiento de datos multivías para la modelación de los datos de las narices electrónicas en un entorno superior a un arreglo bidimensional, abre un campo grande de trabajo para el mejoramiento del rendimiento de los sensores y del proceso de clasificación en estos sistemas.
- Las perspectivas de introducción al enfoque lógico combinatorio en el quehacer quimiométrico, debe ser un campo de acción para el desarrollo de nuevas investigaciones, sobre todo en el campo de la industria del papel y su conservación, atendiendo a la presencia de datos mezclados.

El desarrollo de equipos que combinan la espectroscopia molecular con el análisis de imágenes químicas asociadas a la respuesta espectral, brinda un nuevo enfoque en el análisis de imágenes e incrementa las posibilidades en cuanto a la extracción de información, relevante al complementarse los datos químicos estructurales de la molécula con los que aporta la imagen del compuesto o producto relacionado al desarrollo de métodos y herramientas que permita el descubrimiento, caracterización y visualización 3D en estas relaciones y vínculos. Es un campo abierto a las investigaciones.

4 Principales grupos de investigación en Quimiometría a nivel Mundial

IBEROAMÉRICA:

ESPAÑA.

Universidad de Barcelona

Líneas de investigación:

- Desarrollo de métodos Quimiométricos
- Soluciones de equilibrio de moléculas biológicas.
- Métodos de análisis y automatización de estudios de medio ambiente.
- Desarrollo de sensores químicos.
- Calibración multivariable.

Principales Personalidades:

- Dr. Romà Tauler Ferré

Investigadores principales:

- Dra. Anna de Juan.
- Dr. Raimundo Gargallo

Dirección: www.ub.es/gesq/eq1_cat.htm

Universidad de Burgos

Líneas de investigación:

- Calibración multivariante.
- Regresión robusta
- Análisis de datos estadísticos.
- Reconocimiento de Patrones.
- Diseño experimental
- Métodos de validación y estandarización.
- Análisis medioambiental.
- Principales personalidades:
- Maria Cruz Ortiz Fernández

Investigadores principales:

- Dr. Luis A. Sarabia Peinador
- Dra. Ana Herrero Gutiérrez
- Dra. M^a Sagrario Sánchez Pastor
- Dr. Tomás Pérez Pérez

Dirección: www.ubu.es/investig/grupos/eio-1.htm

Universidad de la Coruña

Líneas de investigación:

- Estudios medioambientales
- Aplicaciones en el control de la calidad
- Calibración multivariable
- Principales personalidades:
- Dr. Jose M. Andrade Garda

Dirección: www.udc.es

Universidad de Granada

Líneas de investigación:

- Protocolos de validación de procedimientos analíticos
- Cálculos de incertidumbres
- Optimización
- Parámetros analíticos de Calidad
- Metodologías estadísticas para el Aseguramiento de la Calidad

Principales personalidades:

- Dr. Luis Cuadros Rodríguez

Dirección: www.ugr.es/~ecugr

Universidad de Santiago de Compostela

Líneas de investigación:

- Optimización en cromatografía.
- Simulación
- Algoritmos genéticos
- Diseño experimental
- Matrices super saturadas
- Aplicación de redes neuronales para el reconocimiento de patrones (narices electrónicas)

Principales personalidades:

- Dr. Rafael Cela

Dirección: www.usc.es/gcqprega

Universidad de Valencia

Líneas de investigación:

- Electroforesis capilar y quimiometría
- Modelado y optimización de HPLC
- Principales personalidades:
- Dr. Guillermo Ramis

Dirección: www.uv.es/~hpcehplc

Universidad de Valladolid

Líneas de investigación:

- Aplicación de métodos multivariados en estudios medioambientales
- Optimización de métodos y procedimientos analíticos
- Modelación de reacciones electroquímicas en sales fundidas

Principales personalidades:

- Dr. Rafael Pardo Almudí

Dirección: www.uva.es

BRASIL

Universidad Federal de Paraíba, UFPB

Grupo de Instrumentación y Automatización en Química Analítica y Quimiometría

Líneas de investigación:

- Métodos Quimiométricos en Química Analítica
- Análisis de multicomponentes simultáneos
- Calibración multivariada
- Optimización experimental

Personalidades principales:

- Mario Cesar Ugulino de Araujo
- Investigadores principales
- Teresa Cristina Becerra Saldaña

Dirección: www.quimica.ufpb.br/~laqa

Universidad Estadual de Campinas (UNICAMP)

Laboratorio de Quimiometría teórica y aplicada

Líneas de investigación:

- Análisis de componentes principales(PCA)
- Análisis exploratorios, QSPR, QSAR
- Análisis de componentes de uno o dos métodos- 1,2MCA
- Resolución de picos, análisis exploratorio
- Análisis de clusters jerárquicos
- SIMCA
- Reconocimiento de patrones
- Regresión de componentes principales –PCR
- Calibración multivariada, QSPR-QSAR
- Regresión de mínimos cuadrados parciales
- PARAFAC
- GRAM
- NBRA
- RBL

Personalidad principal:

- Prof. Dra. Márcia M.C. Ferreira

Investigadores principales:

- André Krell
- Cristiano Gomes de Faria
- Edilton de Sousa Barcellos
- Eduardo Escobar
- Fabiana Alves de Lima Ribeiro
- Luciana Cunha Sabino
- Marcelo Antônio Morgano
- Marcos Gutiérrez
- Miriam da Silva Costa
- Reinaldo Francisco Teófilo
- Rudolf Kiralj

Dirección: www.laqa.iqm.unicamp.br

Universidad Estadual de Londrina – UEL

Líneas de investigación

- Utilización de métodos quimiométricos para análisis de multicomponentes.
- Calibración multivariada
- Sistemas complejos
- Programas computacionales
- Análisis de medicamentos

Personalidad principal:

- Ieda Spacino Scriminio

Principales investigadores:

- Dilson Norio Ishikawa
- Suely Rodriguez Cabeleira

Dirección: www.qui.uel.br

Universidad Federal de Goias - UFG

Líneas de investigación:

- Métodos de reconocimiento de patrones
- Regla de los vecinos mas cercanos
- Análisis de componentes principales
- Matriz de datos multivariados
- Optimización de experimentos

Personalidad principal:

- Paolo Sergio de Souza

Investigadores principales:

- Pedro Enrique Ferri

Dirección: www.cnpq.br

AMÉRICA DEL NORTE

CANADÁ

Dalhousie University, Canada; Halifax, Nova Scotia.

Chemometrics Analysis Research Centre

Líneas de investigación:

- Quimiometría como herramienta para datos multivariadas aún no tratadas
- Modelo de medición de errores de covarianza en instrumentación multicanal
- Máxima probabilidad en análisis de factorial paralelo
- Comparación de regresión de componentes principales y de regresión de mínimos cuadrados parciales a través de simulación genérica de mezclas complejas
- Calibración multivariada

Personalidad principal:

- Dr. Peter D. Wentzell (Professor)

Investigadores principales:

- Tobias Karakach (Graduate student)
- Lorenzo Vega Montoto (Graduate student)
- Marc Léger (Graduate student)

Dirección: www.dal.ca/~pdwentze/index.html

ESTADOS UNIDOS

Universidad de South Carolina

Líneas de Investigación:

- Aplicación de la química analítica en la ciencia forense
- Forensic micro espectroscopia forense
- Optimización y diseño experimental y validación de métodos en química analítica
- Estadísticas multivariadas
- Reconocimiento de patrones para la interpretación de datos químicas

Personalidad principal:

- Dr. Stephen L. Morgan, Professor

Dirección: www.sc.edu/faculty/morgan/index.html

Universidad de Cornell, New York

Department of Food Science and Technology, 129 Food Research Laboratory, Geneva, NY

Línea de investigación:

- Quimiometria y Análisis multivariable en la ciencia del alimento

Personalidad principal:

- Profesor Karl J. Siebert

Dirección: www.nysaes.cornell.edu/fst/faculty/siebert

Universidad de Ohio Departamento de Química

Líneas de investigación:

- Desarrollo de algoritmos para el procesamiento, modelado e interpretación de señales en tiempo real
Acoplamiento de métodos Quimiométricos con espectrómetros de masa .
- Diseño de instrumentos químicos inteligentes
- Análisis forenses y biológicos mediante el uso de espectrometría ion móvil

Personalidad principal:

- Profesor Howard D. Dewald
- Profesor Peter de B. Harrington

Dirección: www.chem.ohiou.edu/analytical/cici.html

Universidad de Delaware

Departamento de química y bioquímica

Líneas de investigación:

- Improving the Robustness and Transfer of Multivariate Calibrations
- Data and Model Fusion for Multivariate Classification and Calibration
- Novel Methods for Improved Classification of High-Dimensional Data Sets

Personalidad principal:

- Prof. Steven D. Brown

Investigadores principales:

- Dr. Yang Liu
- Robert N. Feudale
- Anthony Myles

Dirección: www.udel.edu/chemo

EUROPA

Universidad de Umea, Suecia

Líneas de investigación:

- Desarrollo de métodos Quimiométricos
- Química combinatoria
- QSAR
- Procesos en química analítica
- Análisis de datos
- Personalidad principal:
- Profesor Svante Wold
- Profesor Michael Sjöström

Investigadores:

- Associate Professor Paul Geladi
- Ph.D Henrik Antti
- Ph.D Anders Berglund
- Andersson Per M, ,
- Edman Maria,
- Elg Kristina,
- Gabrielsson Jon,
- Linusson Anna,
- Nilsson David,
- Trygg Johan,
- Uppgard Lise-Lott,
- Wikström Malin

Dirección: www.chem.umu.se/dep/ok/research/chemometrics/index.html

Universidad de Telemark, Noruega

Líneas de investigación:

- Quimiometria Acústica
- Análisis de imagen multivariada (MIA)
- Tecnología de sensores
- AMT
- Coordinaciones Paralelas
- Validación y calibración multivariada

Personalidad principal:

- Maths Halstensen

Investigadores principales:

- Kirm Esbensen
- Rolf Ergon:
- Morten Borg:

Dirección: www.kj.uib.no/html/areachemometrics.html

Universidad de Bristol, Inglaterra

Bristol Chemometrics –

Líneas de investigación:

- Pattern Recognition,
- Pattern recognition in pyrolysis GCMS
- Forensic Work, LCNMR and LCMS
- Reaction Monitoring and Coupled Chromatography.
- LC-NMR and LC-MS
- Biological aspects of chemometrics
- Chemometric data analysis for cinetica

Personalidad principal:

- Richard Brereton

Investigadores principales:

- Tom Thurston
- Sukri Hassan
- Antonio Carvalho
- Mohammed Wasim
- Jirut Wattoo
- Simeone Zomer
- Lifeng Zhu
- M.Sc. by Research Students
- Yun Xu

Dirección: www.chem.bris.ac.uk/org/chemometrics

Universidad de Nijmegen, Belgium

Laboratory of Analytical Chemistry (LAC)

Líneas de investigación:

- Quimiometria metodológica
- Estadísticas multivariantes
- Análisis de imágenes espectroscópicas.
- Spectroscopic image analysis: linking pixel-based quantification or classification techniques to image processing techniques.
- Quimiometria molecular

Personalidad principal:

- Prof. Dr. L.M.C. Buydens

Dirección: www.cac.sci.kun.nl

Universidad de Viena
Instituto de Química Analítica

Líneas de investigación:

- Aplicación de métodos quimiométricos a la cromatografía y a la espectrometría de masa
- Desarrollo de programas quimiométricos y la promoción para su uso
- Deconvolución de picos mediante el uso de análisis factorial
- Desarrollo de un algoritmo para el reconocimiento de picos basado en técnicas de reconocimiento de patrones
- Investigación sistemática de las posibilidades de reconocimiento de subestructuras de bibliotecas (libraries) de espectrometría de masa de baja resolución

Personalidad principal:

- Dr. Gregor Reich

Dirección: www.anc.univie.at/chemomrt.html

Duke University Department of Chemistry

Líneas de investigación:

- Análisis quimiométrico de datos de retención
- Principales componentes de análisis factorial(factor)
- Optimización de la separación mediante HPLC
- Investigación de la respuesta cromatográfica como función de detección de onda
- Análisis cuantitativos de multicomponentes fluorescente
- Mezclas utilizando análisis quimiométricos

- Análisis quimiométrico de datos multivariantes
- PCFA
- TTFA

Personalidad principal:

- Profesor Charles H. Lochmüller

Dirección: www.chem.duke.edu/~clomul/indexn.htm

Sociedades quimiométricas más importantes

- Sociedad quimiométrica de Bélgica
- Sociedad quimiométrica de Dinamarca
- Sociedad quimiométrica de Holanda
- Sociedad quimiométrica de Francia
- Sociedad quimiométrica de Alemania
- Sociedad quimiométrica de la Gran Bretaña
- Sociedad quimiométrica de Italia
- Sociedad quimiométrica de Noruega
- Sociedad quimiométrica de Rusia
- Sociedad quimiométrica de España
- Sociedad quimiométrica de Portugal
- Sociedad quimiométrica de Suecia
- Sociedad quimiométrica Internacional
- Sociedad quimiométrica de Estados Unidos

5 Publicaciones

Revistas de impacto

- The Analyst . Analytical Chemistry .
- Analytica Chimica Acta)
- Chemistry in Britain .
- Chemometrics and Intelligent Laboratory Systems .
- Journal of Chemical Information and Computer Sciences.
- Journal of Chemometrics .
- Journal of Chromatography A .
- Rapid Communications in Mass Spectrometry .
- Trends in Analytical Chemistry .

Sitios de Quimiometría

- Wiley Chemometrics World – nuevos artículos, recursos, vínculos y más. Incluye el sitio Spectroscopy Now
- Chemometrics Online – Sitio que ofrece análisis en línea, Análisis de datos. Soporta PCA, PCR, PLS, SIMCA y wavelet para la eliminación de ruidos en datos.
- Chemometrics Web News – Un sitio relativamente nuevo, soportado por el grupo de Quimiometría de la Universidad de Milán.
- Homepage of Chemometrics - Por Johan Trygg de la Universidad Umeå. Noruega.
- NAMICS - homepage de la North American Chapter of the International Chemometrics Society.
- ChemoBro Literature Database –Base de datos para la búsqueda de publicaciones de Quimiometría.
- Chemometrics.net – noticias & Vínculos.
- An Introduction to Chemometrics – Sitio Web que contiene una colección de slides para el estudio introductorio en Quimiometría.
- Chemometrics - Online lecture notes for introductory courses to many chemometric techniques, include ANOVA, experimental design, PCA and calibration (University of Akron)

Libros

- Chemometrics, by M.A. Sharaf, D.L. Illman, and B.R. Kowalski.
- Multivariate Calibration, by H. Martens and T. Naes.
- Chemometrics: A Practical Guide, by K.R. Beebe, R.J. Pell, and M.B. Seasholtz.
- Handbook of Chemometrics and Qualimetrics, by D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, and J. Smeyers-Verbeke.
- A User's Guide to Principal Components, by J. Edward Jackson
- Multivariate Data Analysis In Practice. 5th ed.
ISBN 82-993330-3-2

- 1



Chemometric Techniques for Quantitative Analysis
Richard Kramer
215 pages, illustrated printed on acid-free paper
Marcel-Dekker (1998) ISBN 0-8247-0198-4 \$99.75
- 2



Data Analysis for the Laboratory and Chemical Plant
Richard Brereton
504 pages, illustrated hard cover
John Wiley (2003) ISBN 0-471-48977-8 \$204.00
paperback ISBN 0-471-48978-6 \$79.00
- 3



Factor Analysis in Chemistry, Third Edition
Edmund Malinowski
432 pages, illustrated printed on acid-free paper
John Wiley (2002) ISBN 0-471-13479-1 \$115.00
- 4



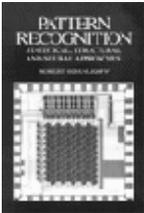
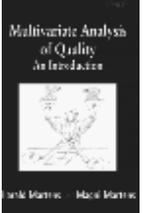
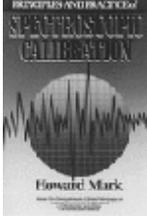
Applied Regression Analysis, Third Edition
Norman Draper, Harry Smith
709 pages, illustrated
John Wiley (1998) ISBN 0-471-17082-8 \$115.00
- 5

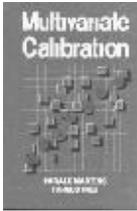
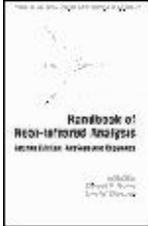
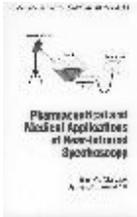


Basic Statistics and Pharmaceutical Statistical Applications
James E. De Muth
596 pages, illustrated printed on acid-free paper
Marcel Dekker (1999) ISBN 0-8247-1967-0 \$79.95
- 6



A User's Guide to Principal Components
J. Edward Jackson
592 pages, illustrated
John Wiley (1991) ISBN 0-471-62267-2 \$198.00
paperback ISBN 0-471-47134-8 \$89.95

- 7  **Pattern Recognition Statistical Structural and Neural Approaches**
Robert Schalkoff
364 pages, illustrated
John Wiley (1992) ISBN 0-471-52974-5 \$106.95
- 8  **Multivariate Analysis of Quality: An Introduction**
Harald Martens, Magni Martens
466 pages, illustrated
John Wiley (2001) ISBN 0-471-97428-5 \$232.00
- 9  **Chemometrics**
Statistics and Computer Application in Analytical Chemistry
Matthias Otto
314 pages, illustrated printed on acid-free paper
paperback
Wiley - VCH (1999) ISBN 3-527-29628-X
\$75.00
- 10  **Chemometrics, A Practical Guide**
Kenneth Beebe, Randy Pell, Mary Beth Seasholtz
348 pages, illustrated printed on acid-free paper
John Wiley (1998) ISBN 0-471-12451-6
\$99.95
- 11  **Principles and Practice of Spectroscopic Calibration**
Howard Mark
192 pages, illustrated paperback
John Wiley (1991) ISBN 0-471-54614-3
\$142.00
- 12  **Chemometrics in Environmental Analysis**
J.W. Einax, H.W. Zwanziger, S. Geiss
404 pages, illustrated
VCH (1997) ISBN 3-527-28772-8 \$185.00

- | | | |
|----|---|--|
| 13 |  | <p>Multivariate Calibration
Harald Martens, Tormod Naes
416 pages, illustrated paperback
John Wiley (1989) ISBN 0-471-93047-4 out of print</p> |
| 14 |  | <p>Handbook of Near-Infrared Analysis
D.A. Burns, E. W. Ciurczak, Editors
820 pages, illustrated
Marcel-Dekker (2001) ISBN 0-8247-0534-3
\$225.00</p> |
| 15 |  | <p>Pharmaceutical and Medical Applications of NIR Spectroscopy
E. W. Ciurczak, J. K. Drennen
208 pages, illustrated
Marcel Dekker (2002) ISBN 0-8247-9453-2
\$135.00</p> |
| 16 |  | <p>Near-Infrared Spectroscopy: Principles, Instruments, Applications
H. Siesler, Y. Ozaki, S. Kawata, H. M. Heise
361 pages, illustrated
John Wiley (2002) ISBN 3-527-30149-6
\$165.00</p> |

Otros libros de interés

- Hyperstat Online Textbook - an excellent textbook by Prof. David Lane, also containing many links to other useful resources
- Introductory Statistics - an online textbook teaching statistics by David Stockburger, suitable for those new to the subject
- Multivariate Statistics - a very comprehensive online textbook for multivariate statistics, again by David Stockburger
- Statsoft Text Book - another good site for learning statistics
- Multivariate Statistics: a Practical Guide - another useful online textbook, this time written by Mike Wulder
- Numerical Recipes - website containing full online versions of the books published by the Cambridge University Press.

6 Principales softwares utilizados en Quimiometría

El análisis de los Softwares objetos de estudio lo hemos centrado precisamente en tres aspectos fundamentales que constituyen problemas esenciales para el diseño de un sistema de RP.

1. Representación de los datos de entrada.
2. Extracción de los atributos y características de los datos de entrada y reducción de la dimensionalidad de los vectores patrones (preprocesamiento y extracción de los rasgos).
3. Determinación de los procedimientos óptimos de decisión en los procesos de clasificación e identificación.

Software	Rep. de los Datos	Ext. Atrib. y Caract.	Proc. de desic.
Pirouette	Numérico-Matricial	Utiliza herramientas estadísticas basadas en los algoritmos PCA y HCA además de Modelos de predicción basados en el CLS y el FBR. En fenómenos rugosos o difusos utiliza los métodos Alternating Least Squares ALS Multivariate Curve Resolution MCR	K Nearest Neighbors (K-NN) Soft Independent Modeling of Class Analogy (SIMCA) Calibration Transfer (CT)
PLS_Toolbox	Numérico-Matricial	Utiliza para el análisis exploratorio de los datos el método PCA y otras herramientas estadísticas como el CLS, MLR, SMLR, PCR, PLS, Generalized Rank Annihilation Method (GRAM), inverse least squares (ILS) model, PARallel FACTor analysis, PARAFAC, como aspecto significativo (PCA para Imágenes)	Soft Independent Modeling of Class Analogy (SIMCA)

Unscrambler	Numérico-Matricial	Utiliza para el análisis multivariado el PCA,	SIMCA, PCR, PLS, Modelos de
-------------	--------------------	---	-----------------------------

		PCR, PLS y MLR, Modelos de Proyección de Variables Latentes, LVRM, ANOVA	Proyección de Variables Latentes, LVRM
SIMCA-P	Numérico-Matricial	Utiliza PCA, PLS, SIMCA, Orthogonal Signal Correction OSC, Wavelets Compression WC, Modelos de distancia	SIMCA

Sirius	Numérico-Matricial	Utiliza PCA, PLS, SIMCA, Factorial Fraccional Design FFD, Modelos de Proyección de Variables Latentes, LVRM, Dendogramas.	PCA, PLS, SIMCA
Extract	Numérico-Matricial	Utiliza el PCA, PCR, PLS	PCA, PLS para la clasificación.
MVA	Numérico-Matricial	Utiliza el PCA, PCR, PLS y MLR	-

Peculiaridades

De los Softwares estudiados, el Pirouette, SIRUS y SIMCA-P estan concebidos con un enfoque mas flexible al reconocimiento de patrones.

El PLS_Toolbox consiste en un paquete de herramientas independientes que trabajan en un ambiente Matlab a partir de cuyas combinaciones y otras herramientas es posible la construcción de Sistemas de RP para resolver problemas de la Quimiometría.

El resto (Unscrambler, Extract y MVA) brindan herramientas de gran fortaleza para la exploración y extracción de la información de los datos.

Todos los softwares analizados ofrecen utilidades gráficas excepto el MVA que de hecho se convierte en una herramienta de Microsoft Excel.

Formatos que soportan y aplicaciones generales

Software	Formatos de entrada	Aplicaciones
Pirouette	Extenciones .PIR, .DAT, .WKS y .XLS	Monitoreo más comprensivo de la calidad de un producto y cambios en los parámetros de un proceso. Monitoreo de la ruta de calidad de materiales incluyendo su origen geográfico y varietal

		<p>Reemplazamiento o argumentación de evaluación sensorial con sistemas de instrumentos analíticos.</p> <p>Detección más eficiente de adulteración de productos, contaminación y sustitución.</p>
PLS_Toolbox	.MAT, .DAT, .XLS	Procesos relacionados con la Ingeniería Química y la Química Analítica.

Unscrambler	.DAT, .WKS, .XLS, .MAT , acepta datos proveniente de software de algunos equipos analíticos (*.asd, .001 etc.)	Reformar, Transformar y Plotear sus datos. Estudiar la cantidad de variaciones de un grupo de variables. Estudiar la relación entre un grupo de variables. Validar sus modelos multivariados con test de incertidumbre. Estimar nuevos valores de respuesta desconocidos. Clasificación de muestras desconocidas. Realizar Diseño Experimental de plantas.
SIMCA-P	.DIF, .TXT, .XLS, .WKL , y datos proveniente de software de algunos equipos analíticos (Cromatografía y .spc)	Clasificación, Predicción en Cromatografía, espectroscopia y otros análisis químicos.
Sirius	.00, .TXT, .XLS, .WKL ,	Exploración de datos multivariados. Clasificación. Exploración y Optimización de procesos por el diseño factorial fraccional.
Extract	.EXD, .DIF, .TXT, .XLS, .DAT, .MAT, .UNS, .INP	Exploración de datos multivariados. Clasificación.
MVA	.XLS	Extracción y exploración de los datos

Principales compañías productoras de Herramientas en Quimiometría:

- Applied Chemometrics
- CAMO – Produce un número de software, entre los que se incluye el Unscrambler.
- Deconvolution and Entropy Consulting - produce software de análisis espectral usando métodos Bayesianos, Máxima Entropía, y otros.

- Eigenvector Research – Productores del PLS_Toolbox para Matlab, que contiene un variado conjunto de métodos de análisis de datos.
- Galactic - Produce un número de software para su uso en espectroscopia. Utiliza un also have a useful algorithm reference section.
- Infometrix - makers of the Pirouette software package.
- Lab Control - creators of Spectacle software for spectroscopic and chromatographic data processing.
- Mathsoft - produce the MathCad environment, and also have a section of wavelet resources.
- MathWorks - the people behind Matlab, a matrix-based mathematical environment.
- Maximum Entropy Data Consultants - specialise in custom and semi-custom software for Bayesian Data Analysis.
- Statistical Designs - makers of the MultiSimplex optimisation software package.
- Three Mode Company - company specialising in three-mode data analysis.
- Trilobyte - provide statistical software, including S-Plus.
- Umetrics - producers of MODDE (experimental design) and SIMCA (data analysis) software.
- Wavebox - provide the WavBox Software Library and FirWav Filter Library for wavelet transforms

Software links

Charm Works Operates in Lab VIEW environment, no demo available.

Evince Chemometric software by UmBio, demo available

Extract Chemometric software by Otto Lindeberg, demo available

Matlab homepage | Matlab toolboxes | More toolboxes |

MINITAB Statistical software, demo available.

Modde Umetrics AB software, demo available.

MultiSimplex Experimental design and optimization.

ParLeS, download freeware that does PLS with leave one out cross validation

PLSplus/IQ software (Galactic Software).

Pirouette Infometrix software, demo available.

PLS-Toolbox Eigenvector Research Inc.

Simca Umetrics AB software, demo available

Sirius, SirExtricate Pattern Recognition Software (PRS), Norway.

Chemometrics Toolbox for MATLAB

Unscrambler CAMO, demo available.

1. Neural Network 4 (D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J.Lewi, J Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier Science, Amsterdam, 1997.
2. Guillermo Ramis Ramos, M^a.Celia Garcia Álvarez-Coque, Quimiometría, Editorial Síntesis, S.A. España, 2001.
3. Svante Wold, Special Session on the History, Foundations and the Founding Fathers of Chemometrics, Proceedings 9th International Conference CAC 2004, Lisbon, Portugal, 20-23 September 2004.
4. L.A. Currie, J.J.Filliben, J.R DeVoe, Statistical and mathematical method in analytical chemistry, Anal Chem, 44(1972) 497R-512R.
5. B.R.Kowalski, Chemometrics: view and propositions, J. Chem. Inf. Comp. Sci. 15 (1975), 201-203.

6. J.T. Clerc, E Ziegler, Editorial, *Anal. Chim. Acta* 95 (1997).
7. A.M.G. McDonald, H.L.Pardue, A.Townshend, J.T. Clerc, Editorial, *Anal. Chim. Acta.* 134 (1982) 1.
8. Barry K. Lavine, *Chemometrics, Anal.Chem.* (1998) 70, 209R-228R.
9. Barry K. Lavine, *Chemometrics, Anal.Chem.* (2000) 72, 91R-97R.
10. Barry K. Lavine, Jerome Workman, *Chemometrics* (2002), 74, 2763-2770.
11. Richard G. Brereton, *Chemometric, Data Analysis for the Laboratory and Chemical Plant.* John Wiley & Sons, Ltd. 2002.
12. Kim H. Esbensen, *Multivariate Data Analysis in Practice.* 5th Edition Camo Process As, 2002.
13. B.A.Rock, *An Introduction to Chemometrics, General Areas,* 469D, 7- 1119/jun 6 1997.
14. Philip K. Hopke, *The Evolution of Chemometrics Review, Anal.Chem. Acta.* 500 (2003) 365-377.
15. D.L. Massart, L. Kaufman, *Interpretation of Analytical Chemical Data By the Use of Cluster Analysis.* Wiley, New York, (1983).
16. T.Kohonen, *Self –Organization and Associative Memory,* third Ed. Springer Verlag, New York 1989.
17. S.Grossberg. *Adaptive pattern classification and universal recoding. I. Parallel development and coding of neural feature detectors,* *Biol. Cybern.* 23 (1976) 121-134.
18. S. Grossberg. *Adaptive pattern classification and universal recoding II. Feedback, expectation, olfaction, and illusions,* *Biol.Cybern.* 23 (1976) 87-203.
19. G.A.Carpenter, S.Grossberg, D.B. Rosen, *ART-2a-an adaptive Resonance algorithm for rapid category learning and recognition* (1991) 493-504.
20. D.Wienke, L. Bydens, *Adaptive resonance theory based neural Networks- the ART of real- time pattern recognition in chemical process monitoring.* *Trends Anal.Chem.* 14(1995) 398-406.
21. Y.Xie, P.K.Hopke, D. Wienke, *Airborne particle classification with a combination of chemical composition and shape index utilizing an Adaptive resonance artificial neural network,* *Environ.Sci.Technol.* 28, (1994) 1921-1928.
22. X.H. Song, P.K. Hopke, D.P. Fergenson, K.A Prather, *Classification of single particles analyzed by ATOFMS using an artificial neural network, ART-2a.* *Anal. Chem.* 71 (1999) 869-865.
23. N.Cristianini, J.Shawe-Taylor, *An Introduction to Support Vector Machines,* Cambridge University Press, Cambridge, 2000.
24. V. Vapnik, *Statistical Learning Theory,* Wiley Interscience, New York 1998.
25. Zomer.S, *Support Vector machines for the Discrimination of analytical Chemical Data application to the determination of Tablet Production by Pyrolysis-Gas Chromatographic.* *The Analyst* 2004; 129: 175-181
26. N.J.Nilson, *Linear Learning Machines,* Mc Graw-Hill, New York 1965.
27. D.Coomans, I. Broekaert, *Potential Pattern Recognition in Chemical and Medical Decisions Making,* Wiley, New York. 1986.
28. S.Wold. *Pattern recognition by means of disjoint principal components Models,* *Pattern recognition* 8 (1976).
29. M.P.Derde, D.L.Massart, *UNEQ: a disjoint modelling technique for Pattern recognition based in normal distribution.* *Anal.Chim.Acta.* 184 (1986) 33-51.
30. M.P. Derde, L. Bydens, C. Guns, D.L. Massart, P.K. Hopke, *Comparison of Rule Building expert systems with Pattern Recognition For the classification of analytical data,* *Anal. Chem* 59 (1987) 1868-1871.
31. J.R.Quinlan. *Learning efficient classification procedures and their Applications to chess and games in R.S Micahlski, J.G.Carbonell, T.M.Mitchell (Eds.), Machine Learning: An Artificial Intelligence Approach,* Tioga, Palo Alto, CA, 1983, pp 463-482.
32. Wold S, Sjostrom.M *Chemom, Intell.Lab.Syst.* 1998, 44(1, 2), 3-14.
33. Brown, S. D. *Comp.Chem.Eng.* 1998, 23(2), 203-216.
34. Wentzell, P.D, Lohnes.M.T, *Chemom.Intell. Lab.Syst.* 1999, 45(1, 2), 65-85.
35. Grung, B., Manne, R. *Chemom.Intell. Lab.Syst.* 1998, 42(1,2), 125-139.
36. Hopke.P.K, Xie.Y.,Paatero, P, *Journal Chemom.* 1999, 13(3-4),343-352.
37. Rutledge. D, Barros, A. *Theory and Applications of the Principal Components Transform: A New Tool for Data Analysis.* *Proceedings CAC 2004 Lisbon,* Abstract 196 pag. 52.
38. Barros, A, Rutledge. D, *Seg PCT: Segmented principal Component Transform.* *Proceedings CAC 2004 Lisbon,* Abstract 1 pag. 53.
39. Cai.C, de Harrington, *Anal.Chem.* 1999, 71(19), 4134-4141.
40. Jalali-Heravi. Fatemi.M.H, *Journal Chromatography* 1998, 825(2), 161-169.
41. Barden. C.J, *J.Chem. Inf.Comp.Sci.* 1998,38(3), 483-488.
42. Liu. L, *J.Inclusion Phenom.Macrocyclic. Chem.,* 1999, 34-(3), 291-298.
43. Tetteh. *J.Chem. Inf.Comp. Sci.* 1999, 39 (3), 491-507.
44. Hall. L.H, Story, C.T. *SAR QSAR Environ.Res.* 1997, 6 (3- 4), 139-161.

45. Huuskonen, J. *J. Chem. Inf. Comp. Sci.*, 1998, 38(3), 450-456.
46. Wold, S., Norden, B. *Chemom. Intell. Lab. Syst.*, 1998, 44(1, 2), 213-227.
47. Barros, A., Rudledge, D., A novel technique for cluster analysis. *Chemom. Intell. Lab. Syst.*, 70, 2004, 99-112.
48. Barros, A., Rudledge, D. Applications of PLS Cluster to the Selection of Variables. *Proceedings CAC, Lisbon 2004*, abstract 2.
49. Gasteiger, J. *J. Chem. Inf. Comp. Sci.*, 1998, 38(2), 210-219.
50. Meusinger, R. *Chemom. Intell. Lab. Syst.*, 1999, 46(1), 67-68.
51. Lanteri, S., Armanino, C. *Proceedings CAC, Lisbon 2004*, abstract 165.
52. Varmuza, K., Demuth, W. Modelling Gas Chromatographic Retention Indices by a Large Set of General Molecular Descriptors. *Proceedings CAC, Lisbon 2004*, abstract 98.
53. Garkani-Nejad, Varmuza, K., *Chromatogra. A.* 1028, 287 (2004).
54. Todeschini, R., Consonni, V. Software Dragon, versión 4. (2004). <http://dis-ta.unimib.it/chm/Dragon.htm>, Milan, Italia.
55. M. Bos. *Chemom. Intell. Lab. Syst.* 23(1994), 115.
56. B.K. Alsborg, *Analyst* 122(7) 1997, 645.
57. C.R. Mittermayr, *Chemom. Intell. Lab. Syst.*, 34 (1996) 187-202.
58. B. Walczak, D.L. Massart. *Chemom. Intell. Lab. Syst.* 38 (1997) 39.
59. B. Walczak, D.L. Massart. *Anal. Chem.* 68(1996) 1742.
60. U. Depczynski. *Chemom. Intell. Lab. Syst.* 47 (1999) 179.
61. D. Jouan-Rimbaud, B. Walczak. *Anal. Chem.* 69 (1997) 4317.
62. Bakshi, B.R. *AIChE J.* 1996, 42(2) 477-492.
63. Kimble, K.R. Report AEDC-TR-94-17, Order No. ADA289776. Avail. NTIS 1995, 95(19), Abstr. No. 19-03,401.
64. Wolkenstein, M.J. *Trace Microprobe Technol.* 1997, 15(1), 33-39.
65. Kosanovich, K.A. *J. Compt. Chem. Eng.* 1997, 21(6), 601-620.
66. Liu, W. *Chim. Sci. Bull.* 1997, 42(10), 822-825.
67. Neue, G. *Solid State Nucl. Magn. Reson.* 1996, 5(4) 305-314.
68. Chau, F.T. *Appl. Spectrosc.* 1996, 50(3), 339-348.
69. Lu, X.Q. *Analyst* 121 (8), 1029-1024.
70. Rying, E.A. *Proc. Electrochem. Soc.* 1997, 97-9, 37-44.
71. Qi, G. *Compos. Sci. Technol.* 1997, 57(4), 389-403.
72. Susuki, H. *Acoust. Emiss.* 1996, 14(2), 69-84.
73. K. Jetter, U. Depczynski. *Anal. Chem. Acta.* 420 (2000) 169-180
74. Habb, B. *Anal. Chem.* 1995, 67(18), 3253-3260.
75. Cooke, G.J. *Vac. Sci. Technol.* 1996, 14(1) 283-286.
76. Dowset, M.G. *Philos. Trans. R. Soc. London, Ser.* 1996, 354(1719), 2713-2729.
77. Zhang, J. *Am. Soc. Mass. Spectrom.* 1997, 8(6) 659-670.
78. Schmeider, P.J. *Magn. Reson.* 1997, 125(2), 332-339.
79. Cross, K.J. *Data, Handl. Sci. Technol.* 1996 (18), 129-144
80. Riris, H.J. *Proc. SPIE-Int. Soc. Opt. Eng.* 1995, 2366, 90-97.
81. Jiang, J.H., Wang, J.H. *J. Chemom.* 1996, 10(3), 241-252.
82. Dong, D., MacAvoy, J. *Compt. Chem. Eng.* 1996, 20(1), 65.
83. Chen, D. *Comp. Chem. (Oxford)* 1996, 21(2), 109-113.
84. Sagrario Sanchez, *Anal. Chim. Acta.* 1997, 348(1-3), 533-542.
85. Cleij, P. *Anal. Chim. Acta.* 1997, 348(1-3), 495-501.
86. Salenics, A.J. *Process Control* 1997, 7(3), 155-167.
87. Jiang, J. *J. Chemom.* 1996, 10(4), 281-294.
88. Wu, W. *Anal. Chim. Acta.* 1996, 335(1-2), 11-22.
89. Li, Y. *J. Chemom.* 1999, 13(1), 3-13.
90. Carrieri, A.H., *Appl. Opt.* 1999, 38(17), 3759-3766.
91. Lavine, B.K. *J. Anal. Appl. Pyrolysis* 1999, 50(1), 47-62.
92. Lavine, B.K. *Microchem. J.* 1999, 61(1), 33-45.
93. Marini, F., Zupan, J., on the use of a Kohonen neural network as a Class-Modeling Tool. *Proceedings CAC 2004 Lisbon*, Abstract 42 pag. 58.
94. Zhang, M.H., D.L. Massart. Boosting Classification and Regression Trees, *Proceedings CAC 2004 Lisbon*,

- Abstract125, pag.60.
95. Zomer.S. Classification with Support Vector Machines. Homepage of Chemometrics.Nov 2004.
 96. Ryman. Tubb. Proc.SPIE-Int.Soc.Opt.Eng.1996, 2878, 117-127 (virtual Intelligence), 117-127.
 97. Wang.P, J.Sens.Acuaators.1996.B37 (3), 169-174.
 98. Ping.W.Meas.Science.Technol.1996, 7(12), 1707-1712.
 99. Avaritsiotis. J.Sens.Acuaators.B1996, B33 (1-3) ,77-82.
 100. Di Natale, J.Sens.Acuaators. B 1995, B (1-3) ,808-812.
 101. Menzel.R J.Sens.Acuaators.B2000, B68 (1-3) ,115-122.
 102. Polikar.R, Porter.M. Sens.Acuaators.B2001, B80 (3) 243-254.
 103. Akbaryan, F. J.Compt. Chem.Eng.2001, 25(9-10) 1313-1339.
 104. Doynaz. Chemom.Intell. Lab.Syst.2001, 55,(1-2),109-123.
 105. Skov.T. New method for handling Electronic Nose Data. Proceedings CAC 2004 Lisbon, Abstract.32, pag.67.
 106. Schweitzer.R.J.Managing.Mod.Lab.2000,5(1),7-13.
 107. Egelanddal.B. Wold Scanning 1999, 21(5), 316-325.
 108. Thorbjorn.T, Geladi.P, J.Chemom. 2000, 14(5-6), 585-598.
 109. Buydens.L.M.C, Chemical and Diagnostic Information from Magnetic. Resonance Spectra and Images. Proceedings CAC 2004 Lisbon, Abstract. 48. Pag. 39.
 110. Nicholson. Prog.Nucl.Magn.Reson.Spectrosc.2001, 39, (1) 1-40.
 111. Windig.W. Chemom.Intell. Lab.Syst, 1999, 46 (2) 207-219.
 112. Windig.W. J.Chemom.1999, 13 (2) 95-110
 113. Ferreira.M. Quantitative analysis and Classifications of AFM images Of Human Hair. Proceedings CAC 2004 Lisbon Abstract. 186. Pag.41.
 114. Huang. J, Esbensen.K.H, Chemom.Intell. Lab.Syst. 2001, 57,(1)37-57.
 115. Huang. J, Esbensen.K.H, Chemom.Intell. Lab.Syst.2000, 33(7), 471-482.
 116. Kaczmarek. K, Walczak. Enhancement of Images from 2-D Gel Electrophoresis. Proceedings CAC 2004 Lisbon Abstract 75 Pag.171
 117. Wise, B. Introduction to Multivariate Image Analysis, Workshop, CAC 2004 Lisbon.

Anexo 1: Bibliografía de consulta sobre investigaciones aplicadas novedosas desarrolladas en el período en el área del Reconocimiento de Patrones en la Quimiometría.

1. Kemsley, E. K; Ruault. S.; Wilson, R H. Food Chem. 1995, 54(3), 321-326.
2. Briandet. R, Kemsley. E. K. Wilson.R H.j. Agric. Food Chem.1996, 44(1), 170-174.
3. White.D. R..Jr. Colloq. Sci. Int. Cafe. [CR]. I. 16th 1995,259-266.
4. Suchanek, M. Filipova, H.: Volka, K; Delgadillo, I.: Davies, A. N. Fresenius. Anal. Chern 1996,354(3).327-332.
5. Pablos, F. González, A G. Talanta 1996,43(3), 415-419.
6. Goodacre, R Appl. Spectrosc. 1997,51(8),1144-1153.
7. Fresno, J. M.; Prieto, B.; Urdiales, R; Sarmiento, R M. Carballo.,Sci. FoodAgric. 1995, 69(3), 339-345.
8. Dewettinck, K.; Dierckx, S., Eichwalder, P., Huyghebaert. A.Lait 1997, 77 (1), 77-89.
9. Vogels, J T. W. E.; Terwel, L., Tas, A C: van den Berg, F.,Dukel. F.: van der Greef, J. I Agric. Food Chern. 1996, 44(1), 175-180.
10. Dettmar, H. P.; Barbour, G. S.; Blackwell, K T.: Vogl, T. P., Alkon.D L: Fry, F. S., Jr.; Totah, J. E., Chambers, T L. Comput.Chem. 1996, 20(2), 261-266.
11. Shaw, P. E., Moshonas, M. G. Food Sci. Technol. (London)1997,30(5),497-501.
12. Jensen, K; Tygesen, T. K: Kesmir, C., Skovgaard, I. M.; Sondergaard, I.J. Agric. Food Chem. 1997,45(1),158-161.
13. Kinoshita, E.; Ozawa, Y, Aisftima, T J Agn'c. Food Chem. 1997, 45(10), 3753-3759.
14. Lizuka, K.; Aishima, T J Food Sci. 1997,62(1), 101-104.
15. Kwon, Y. K: Cho, R. K; Yasumoto', S. Near Infrared Spectrosc Future Waves, Proc. 7th Int. ConL Near Infrared Spectrosc: Meeting Date 1995: Davies, AM. C, Williams, P. C., Eds.: NIR Publications: Chichester, U.K, 1996: pp 422-425.
16. García-López, C: Grane-Teruel,: Berenguer-Navarro, Garcia-Garcia, J. E.; Martín-Carratalá, M. L.J. Agric. Food Chem.1996,44(7), 1751-1755.
17. Plats-Moya, S.: Grane: Teruel, N.; Berenguer-Navarro: V.; Martín Carratalá. M. L. Agric. Food Chem. 1997,45(6), 2093- 2097.
18. Angerosa, F.: Di Giacinto, L.; Vito. R: Cumitini. S. j. Sci. Food Agnc. 1996, 72 (3), 323-328.

19. Aparicio, R.; Morales, M. T.; Alonso, M. V. *J. Am. Oil Chem. Soc.* 1996, 73(10), 253-1264
20. Husain, S.; Devi, K. S.; Krishna, D.; Reddy, P. J. *Chemom. Intell. Lab Syst.* 1996, 35(1), 117-126.
21. Ruberto, G.; Biondi, D.; Rapisarda, P.; Renda, A.; Starrantino, A. *J. Agric. Food Chem.* 1997, 45(8), 3206-3210.
22. Salgueiro, L. R.; Da Cunha, A. P.; Tomas, X.; Canigual, S.; Adzet, T.; Vila, R. *Flavour Fragrance J.* 1997, 12(2), 117-122.
23. Saez, F. *Biochem. Syst. Ecol.* 1995, 23(4), 431-438.
24. Dunlop, P. J.; Bignell, C.M.; Hibbert, D. B. *Aust. J. Bot.* 1997, 45(1), 1-13.
25. Sun, L.; Danzer, K.; Thiel, G. *Fresenius Anal. Chem.* 1997, 359(2), 143-149.
26. Mangas, I. J.; Rodriguez, R.; Moreno, J.; Suarez, B.; Blanco, D. *J. Agric. Food Chem.* 1997, 45(10), 4076-4079.
27. Kim, K. R.; Kim, J. H.; Cheong, E. J.; Jeong, C.-M. *J. Chromatogr. A* 1996, 722(1+2), 303-309.
28. Kim, K. R.; Kim, J. H.; Cheong, E. J. *Int. Symp. Chromatogr., 33rd Anniv. Res. Group Liq. Chromatogr. Jpn.*; Hatano, H.; Hanai, T., Eds., World Scientific: Singapore, 1995; pp 453-457.
29. González-Vinas, M. A.; Pérez-Coello, M. S.; Salvador, M. D.; Cabezudo, M. D.; Martín-Álvarez, P. *J. Food Chem.* 1996, 56(4), 399-403.
30. Ferreira, V.; Fernandez, P.; Cacho, J. F. *J. Food Sci. Technol. (London)* 1996, 29(3), 251-259.
31. Hahn, P.; Smith, L. C. P.; Leboldus, L.; Littman, C.; Somorjai, R. L.; Bezabeh, T. *Cancer Res.* 1997, 57(16), 3398-3401.
32. Preul, M. C.; Caramanos, Z.; Collins, D. L.; Villemure, J.G.; Leblanc, R.; Olivier, A.; Pokrupa, R.; Arnold, D. L. *Nat. Med. (N. Y)* 1996, 2(3), 323-325.
33. de Cetaines, J. D.; Nadal, L.; Leray, G.; Serrai, H.; Lewa, C. *J. Anticancer Res.* 1996, 6, 6, 1451-1460 (3B, Proceedings of the Special Symposium on "Lipid Metabolism and Function in Cancer", 1995).
34. Tate, A. R.; Crabb, S.; Griffiths, J. R.; Howells, S. L.; Mazucco, R. A.; Rodríguez, L. M.; Watson, D. *Anticancer Res.* 1996, 16, 1575-1579.
35. Proceedings of the Special Symposium on "Lipid Metabolism and Function in Cancer", 1995).
36. Benninghoff, L.; von Czarnowski, D.; Denkhäus, E.; Lemke, K. *Spectrochim. Acta. Part B* 1997, 52B(7), 1039-1046.
37. Mains, D. C.; Polissar, N. L.; Günselmal, S. J. *Proc. Natl. Acad. Sci. U.S.A.* 1997, 94(1), 259-264
38. Ng, K.H.; Ong, S.H.; Bradley, D. A.; Looi, L.-M. *Appl. Radiat. Isot.* 1997, 48(1), 104-109.
39. Voorhees, K. J.; Basile, F.; Beverly, M. B.; Abbas-Hawks, C.; Hendricker, A.; Cody, R. B.; Hadfield, T. L. *J. Anal. Appl. Pyrolysis* 1997, 40, 41, 111-134.
40. Goodacre, R.; Hiom, S. J.; Cheeseman, S. L.; Murdoch, D.; Weightman, A. J.; Wade, W. G. *Curr. Microbiol.* 1996, 32(2), 77-81.
41. Goodacre, R.; Howell, S. A.; Noble, W. C.; Neal, M. J. *Zentralblatt Bakteriologie* 1996, 284(4), 501-515.
42. Nilsson, T.; Bassani, M. R.; Larsen, T.O.; Montanarella, L. *J. Mass Spectrom.* 1996, 31(12), 1422-1428.
43. Gleser, L. *J. Chemom. Intell. Lab. Syst.* 1997, 37(1), 15-22.
44. Song, X.-H.; Hopke, P. K. *Environ. Sci. Technol.* 1996, 30(2), 531-535.
45. Powers, S. E.; Villaume, J. F.; Ripp, J. A. *Groundwater Monit. Rem.* 1997, 17(2), 130-140.
46. Ravichandran, S.; Ramanibai, R.; Pundarikanthan, N. V. *J. Hydrol. (Amsterdam)* 1996, 178(1-4), 257-276.
47. Huntley, S. L.; Iannuzzi, T. J.; Avantage, J. D.; Carlson-Lynch, H.; Schmelz, C. W.; Finley, B. L. *Chemosphere* 1997, 34(2), 233-250.
48. Troiano, J.; Nordmark, C.; Barry, T.; Johnson, B. *Environ. Monit. Assess.* 1997, 45(3), 301-318.
49. Mujunen, S.P.; Minkkinen, P.; Holmbom, B.; Oikari, A. *J. Chemom.* 1996, 10(5&6), 411-424.
50. Jones, J. M.; Davies, T. D.; Dorling, S. R. *Water, Air, Soil Pollut.* 1995, 85(3), 1569-1574.
51. Ludvigsen, L.; Albrechtsen, H.-J.; Holst, H.; Christensen, T. H. *FEMS Microbiol. Rev.* 1997, 20(3-4), 447-460.
52. Collantes, E. R.; Duta, R.; Welsh, W. J.; Zielinski, W. L.; Brower, J. *J. Anal. Chem.* 1997, 69(7), 1392-1397.
53. Welsh, W. J.; Lin, W.; Tersigni, S. H.; Collantes, E.; Duta, R.; Carey, M. S.; Zielinski, W. L.; Brower, J.; Spencer, J. A.; Layloff, T. P. *J. Anal. Chem.* 1996, 68(19), 3473-3482.
54. Rosen, J. *Key Eng. Mater.* 1997, 132-136, 1483-1486 Pt. 2. *Euro Ceramics (190)* Remola, J. A.; Lozano, J.; Ruisanchez, I.; Larrechi, M. S.; Rius, F. X.; Zupan.
55. *J. TrAC, Trends Anal. Chem.* 1996, 15(3), 137-151.
56. Hall, M.; Yablonsky, L. *Archaeometry* 1997, 39(2), 369-377.
57. Liebich, V. *Fresenius' J. Anal. Chem.* 1995, 352(5), 420-425.
58. Dolmatova, L.; Ruckebusch, C.; Dupuy, N.; Huvenne, J.-P.; Legrand, P. *Chemom. Intell. Lab. Syst.* 1997, 16(2), 125-140.
59. Fredline, V.; Kokot, S.; Gilbert, C. *Mikrochim. Acta* 1997, 14 (Suppl.), 183-184
60. (Progressive Fourier Transform Spectroscopy). *Chem. Commun.*; Brown, C. W.; Bide, M. J. *Soc. Dyers Colour.* 1997, 113(2), 51-56.

61. Kokot, S.: Stewart. S. Text. Res. j. 1995. 65(11). 643-651

Tabla de Contenido

1	Introducción	2
1.2	Definición y evolución histórica de la quimiometría	2
2	Principales áreas de trabajo	4
2.1	Reconocimiento de Patrones	4
2.1.1	Análisis Exploratorio de Datos	5
2.1.2	Métodos de Clasificación	6
3	Principales investigaciones llevadas a cabo en el área del reconocimiento de	17
4	Principales grupos de investigación en Quimiometría a nivel Mundial	27
5	Publicaciones	37
6	Principales softwares utilizados en quimiometría	42
Anexo 1:	48

RT_002, Julio 2008

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2008

Editor: Lic. Margarita Ilisástigui Avilés

Diseño de Portada: DCG Matilde Galindo Sánchez

RNPS No. 2142

ISSN Solicitado

Indicaciones para los Autores:

Seguir la plantilla que aparece en www.cenatav.co.cu

C E N A T A V

7ma. No. 21812 e/218 y 222, Rpto. Siboney, Playa;

Ciudad de La Habana. Cuba. C.P. 12200

Impreso en Cuba

