

# Análisis Estadístico Multivariado: una Herramienta Estratégica para el Control de Procesos y Calidad en la Industria Agroalimentaria

*Multivariate Statistical Analysis: a Strategic Tool for Quality and Processes Control in Food Industry*

**Zuluaga Dominguez Carlos Mario<sup>1</sup>**

<sup>1</sup> Escuela de Ciencias Básicas, Tecnología e Ingeniería, Universidad Nacional Abierta y a Distancia – UNAD. Bogotá D.C., Colombia. carlos.zuluaga@unad.edu.co

---

## Resumen

El uso de técnicas de estadística multivariada para el control de procesos y calidad en la industria agroalimentaria ha venido en crecimiento significativo desde la mitad de los años setenta, como una consecuencia de la revolución informática que facilitó el análisis de matrices de datos de gran tamaño. A diferencia de los métodos univariados de exploración de datos, la estadística multivariada utiliza como gran pilar el análisis de información descrita por tres o más variables que pueden ser estudiadas simultáneamente y comprendidas de una manera rápida, eficiente y sencilla. Debido al extraordinario avance en las máquinas informáticas, hoy es posible aplicar estas metodologías para resolver problemas extremadamente complejos. Este artículo presenta las técnicas de estadística multivariada más reconocidas, así como la compilación de algunos trabajos que sirven como demostración de su aplicabilidad en el campo de los alimentos.

**Palabras clave:** clasificación, exploración, modelamiento, optimización, regresión

## Abstract

The use of multivariate statistical techniques for quality and process control in the food industry has been growing significantly since the mid-seventies, as a result of the informatics revolution which facilitated the analysis of large data sets. Unlike univariate methods of data exploration, multivariate statistics uses as a major pillar the analysis of information described by three or more variables that can be simultaneously studied and understood in a fast, efficient and easy way. Thanks to the extraordinary advance in computing machines, it is now possible to apply these methodologies to solve extremely complex problems. This article presents the most recognized multivariate statistical techniques, as well as the compilation of some papers that serve as a demonstration of its applicability in the field of foods.

**Keywords:** classification, exploration, modeling, optimization, regression

Recibido: 19/04/2011

Aceptado: 25/05/2011

## Introducción

El control y monitoreo de los procesos en la industria son establecidos generalmente a través de la realización de mediciones de diferente tipo, cuyos datos deben ser analizados posteriormente para la toma de decisiones. Tradicionalmente se utilizan análisis estadísticos sencillos que involucran dos o tres variables simultáneamente (gráficas, diagramas, medias, modas, proporciones, varianzas, covarianzas), pero no se logra aprovechar adecuadamente la totalidad de la información recolectada y se carece de criterios para eliminar aquella que no es relevante.

El análisis multivariado es una técnica que aplica herramientas matemáticas y estadísticas para el estudio de matrices de datos que contienen numerosas variables; su uso ha sido ampliamente difundido y dentro de las aplicaciones de mayor utilidad se encuentra la evaluación de datos relacionados con diferente tipo de características de productos, de manera que se puedan emitir conceptos confiables referentes a la calidad de los mismos o a las condiciones de los procesos involucrados.

Es así como en los últimos años el uso de técnicas estadísticas multivariadas ha alcanzado un reconocimiento en el mundo para el análisis de datos. Particularmente, esta técnica ha tenido una gran utilidad en el control de procesos agroalimentarios así como en la caracterización y tipificación de alimentos de acuerdo con diferentes características como lo pueden ser el origen geográfico, propiedades nutricionales o funcionales, monitoreo de etapas de producción, etc. Este avance también ha sido consecuencia de la evolución de los equipos utilizados en la industria y en los laboratorios de análisis, los cuales entregan una gran cantidad de datos que si no fuera por las técnicas estadísticas sería imposible de analizar de una manera integral.

Esta revisión está por supuesto más allá de ser un completo resumen de la aplicación de técnicas de estadística multivariada para resolver problemas relacionados con la industria agroalimentaria; el objetivo es mostrar los principales campos de acción de esta técnica, junto a la mención de una lista de referencias de la materia. En la mayoría de casos se tratará de evitar entrar en detalle de los métodos matemáticos y de álgebra lineal relacionados a estas técnicas, los cuales si son de interés del lector podrá encontrar en bibliografía específica [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11].

### **Análisis Estadístico Multivariado**

El Análisis Multivariado es una disciplina que utiliza la matemática, estadística y lógica para a) Diseñar o seleccionar procedimientos experimentales óptimos, b) Proveer la máxima información relevante mediante el análisis de datos y, c) Obtener conocimiento de los sistemas bajo análisis. El análisis multivariado de datos es una técnica en la cual se deben estudiar simultáneamente dos o más variables a partir de un número de observaciones [12].

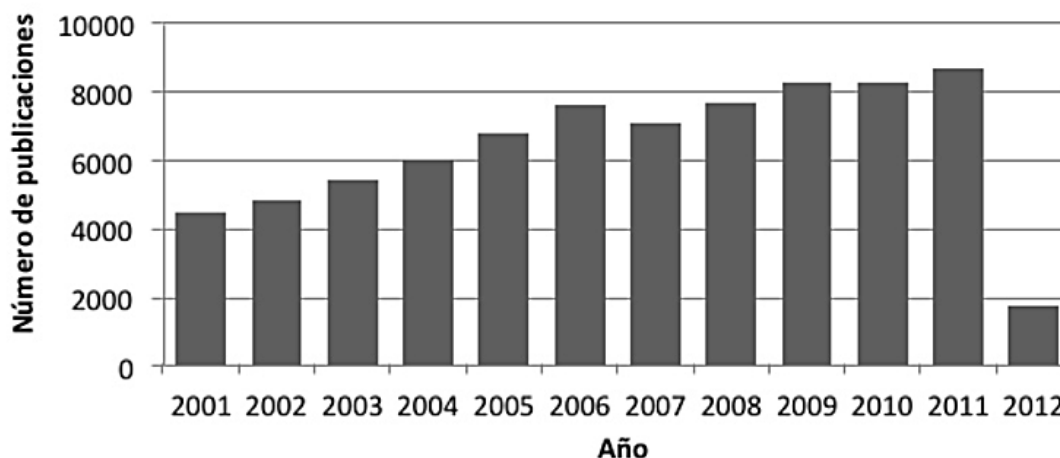
Un análisis multivariado típico parte de un conjunto de datos arreglados en una estructura de dos vías, tabla o matriz. Un ejemplo puede ser una tabla en la cual cada fila corresponda a una muestra (observación) y cada columna a una variable que describa el sistema complejo. Cuando esta matriz es analizada, todas las variables son consideradas al mismo tiempo y, consecuentemente, la información extraída representa una visión global del sistema [13].

Representativamente, se designa como **X** a la matriz de datos en la cual en número de filas (observaciones) y columnas (variables) es usualmente indicado por las letras *n* y *p* respectivamente. En este orden de ideas, una matriz **X** de tamaño 20 x 15 tendrá 20 observaciones a las cuales se le midieron 15 diferentes variables.

### **Análisis Multivariado Aplicado en la Industria Agroalimentaria**

En la actualidad los sistemas de control de producción y las pruebas de autenticidad de alimentos están en creciente demanda en la industria de este sector, ya que representan los temas de interés tanto de consumidores como de productores y comercializadores en el mercado actual. Los sistemas de producción y los productos alimenticios pueden ser descritos como sistemas complejos, en donde varios factores pueden interactuar y jugar un papel fundamental. Consecuentemente, todos estos factores deben ser monitoreados y controlados.

Con base en lo anterior, en los últimos años los reportes bibliográficos sobre el uso de técnicas de estadística multivariada aplicados a la industria agroalimentaria superan los 8000 artículos anuales; este valor supera por el doble el número de publicaciones para esta temática en el año 2000. En la Figura 1 se puede observar el avance en el número de trabajos reportados.



Fuente: Web of knowledge [14]. Thomson Reuters.

**Figura 1. Número de artículos publicados por año sobre estadística multivariada aplicado en la industria agroalimentaria (A marzo de 2012)**

Existe también la necesidad en la industria agroalimentaria de racionalizar y mejorar los procesos de control de calidad. Los sistemas de producción modernos requiere monitoreos rápidos y en línea, los cuales deben ser capaces de extraer la máxima cantidad de información disponible para asegurar un funcionamiento óptimo. Por otra parte, los productos alimenticios adquieren un mayor valor cuando su autenticidad es protegida, controlada y asegurada; de hecho, los consumidores están más orientados hacia adquirir alimentos de un cierto origen geográfico. Es tal la evolución de esta tendencia del mercado que la Comunidad Andina profirió la Decisión 486/2000 que regula en los Países Andinos lo relativo a las Denominaciones de Origen, que en Colombia es controlado por la Resolución 210 del 2001 de la Superintendencia de Industria y Comercio [15], con el fin de garantizar la proveniencia y autenticidad de los productos mediante sellos de calidad.

En resumen, el control de la calidad y autenticidad es un factor que depende de sistemas complejos, descritos por una gran cantidad de datos; como consecuencia, se deben utilizar herramientas específicas para asegurar una interpretación correcta. Una característica fundamental de la estadística multivariada es la simplicidad de sus respuestas: esto significa que los modelos matemáticos usados para interpretar los datos pueden ser complejos y estructurados, pero las respuestas dadas deben ser claras.

Esta revisión está enfocada a la aplicación de la estadística multivariada para el análisis de datos obtenidos en la industria agroalimentaria, y se ha realizado de acuerdo con cuatro métodos estadísticos básicos: exploración de datos, regresión, clasificación y optimización.

### **Métodos de exploración de datos**

Como su nombre lo indica, los métodos de exploración de datos son técnicas con las cuales se busca analizar el comportamiento de las observaciones respecto a todas las variables al mismo tiempo. Dentro de estos métodos, el más reconocido es el Análisis de Componentes Principales (PCA).

El PCA es el método más común usado para presentar la estructura de datos multivariados, es una técnica muy conocida que proyecta los datos en un hiperespacio reducido, definido por los componentes principales; estos son combinaciones lineales de las variables originales, el primer componente principal con la mayor varianza, el segundo componente con la segunda mayor varianza y así sucesivamente. De esta manera, es posible retener un número de componentes menor al número de variables originales,

es decir, es posible reducir la dimensionalidad de los datos: el número de componentes por ser retenido puede ser elegido con base en diferentes parámetros, ligados a la varianza explicada por cada componente principal [1].

Diferentes estudios se han llevado mediante el uso de PCA para el establecimiento de diferentes características en productos alimenticios, estos son presentados en la Tabla 1.

**Tabla 1. Reporte bibliográfico sobre métodos de exploración de datos**

Autor	Estudio	Conclusión
(Baardseth et al., 1995)	Efecto sobre las propiedades sensoriales de productos de pastelería mediante la evaluación de 11 diferentes tipos de grasa de origen animal o vegetal a 3 diferentes concentraciones.	Es posible diferenciar los productos de pastelería según el contenido de grasa y la concentración. Las variables que más influencia tienen son la textura y el color.
(Kallithraka et al., 2001)	Clasificación de 33 vinos griegos de acuerdo con el origen geográfico mediante análisis instrumentales y sensoriales.	Los resultados mostraron que los vinos podían ser divididos en dos zonas geográficas. Las variables que mayor influencia mostraron fueron el contenido de antocianinas y análisis sensorial.
(Rodríguez-Delgado et al., 2002)	Clasificación de vino tinto originario de las Islas Canarias mediante la determinación del contenido de polifenoles por Cromatografía Líquida de Alta Eficiencia.	Es posible diferenciar de acuerdo con la zona de producción los vinos estudiados. La variable de mayor influencia para la clasificación fue el contenido de Quercetina.
(Shin et al., 2010)	Clasificación de cultivos de cacahuete Estadounidenses de acuerdo al contenido de ocho diferentes ácidos grasos.	Se encontró que el ácido oleico es el ácido graso que más diferencia los cultivos. La clasificación se realizó separando los cultivos con alto contenido de ácido oleico con aquellos con bajo contenido.
(Guerreiro et al., 2012)	Evaluación de los factores responsables de la proteólisis del queso en etapa de maduración como marcador de origen geográfico.	Los resultados mostraron que además de los diferentes tipos de caseína presentes en el queso, la fracción proteica de la leche influye fuertemente en las reacciones de proteólisis.
(Patras et al., 2011)	Evaluación de los perfiles antioxidantes de frutas y vegetales en Irlanda	Los resultados de PCA mostraron que los frutos de baya tienen un mayor contenido de fenoles. Algunos vegetales tales como cebolla y brócoli tuvieron una baja capacidad antioxidante.
(Guillén-Casla et al., 2011)	Estudio sobre los efectos en las propiedades de los alimentos al ser irradiados a diferentes dosis.	Radiaciones de 6-8 kGy producen cambios en la composición química de los alimentos, especialmente en aminoácidos, mientras que valores de 1-2 kGy no tienen un efecto sobre estos parámetros.

(De Belie et al., 2000)	Estudio sobre el sonido al masticar manzanas para evaluar la crujencia de acuerdo con su estado de maduración.	Fue posible diferenciar la crujencia de las manzanas de acuerdo a su estado de maduración con base en el sonido generado al masticar. Las frecuencias entre 100–500 Hz contribuyen al clasificar las manzanas crujientes.
(Rodríguez-Campos et al., 2011)	Caracterización de 39 compuestos volátiles y no volátiles generados durante la fermentación de cocoa.	La evaluación por PCA permitió establecer que los compuestos que mayor influencia tienen en la generación de aromas en la fermentación de cacao son el ácido acético y el ácido isobutírico.
(Horimoto et al., 1998)	Clasificación de leche pasteurizada mediante la determinación de off-flavours por cromatografía de gases y panel sensorial.	El análisis estadístico permitió diferenciar muestras normales de aquellas con defectos en aroma y sabor: cocido, recalentado, oxidado.

Fuente: elaboración propia

## Métodos de Clasificación

La clasificación es una de las metodologías fundamentales de la estadística multivariada y consiste en encontrar un modelo matemático capaz de reconocer la pertenencia de un objeto a una clase. Una vez el modelo de clasificación ha sido obtenido, la pertenencia de nuevas observaciones a una clase definida puede ser predicha [16],[17],[20],[63]. En estos casos, adicional a la matriz tradicional de datos, se cuenta con un vector de clases y, el cual asigna la pertenencia a cada observación dada en la matriz, ya sea un origen geográfico o un atributo cualitativo (bueno/malo), (si/no), etc.

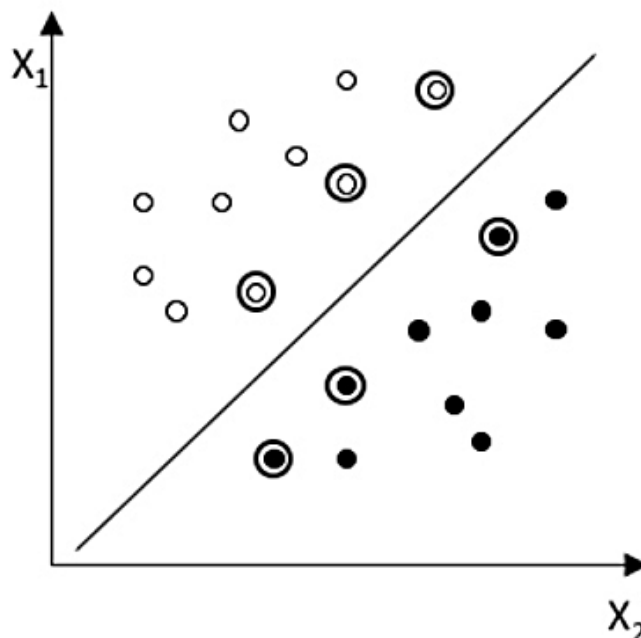
En los métodos de clasificación, el análisis discriminante es uno de los clasificadores más populares. Este método es una técnica de clasificación paramétrica probabilística pues maximiza la varianza entre categorías y minimiza la varianza dentro de las categorías, por medio de la proyección de datos de un espacio de alta dimensionalidad a uno de baja. De esta manera, un número de funciones discriminantes lineales ortogonales igual al número de categorías menos 1 son obtenidas [4].

El Análisis Discriminante Lineal (LDA) es usado dependiendo de la separabilidad de las clases según linealidad/no linealidad y sobre la confiabilidad de las matrices de covarianza de las clases. Para estimar la matriz de covarianza de la clase, el número de objetos de la misma debe ser mayor que el número de variables, mientras que el LDA puede ser aplicado sólo si el número total de muestras es mayor que el número de variables [6].

La relación representa la manera como la descripción del sistema concuerda con la clase y esta se asocia a la calidad descriptiva del modelo (también conocido como *fitting*); entre tanto, si el modelo es planeado para usos futuros, la calidad predictiva del modelo debe ser verificada, conocido como validación [9].

La validación es un conjunto de procedimientos realizados para estimar la habilidad predictiva del modelo. Busca la complejidad óptima lo que evita modelos donde el desajuste de esta capacidad sea debida a modelos muy sencillos (alto sesgo) por falta de información útil o modelos muy complejos (alta varianza) por la presencia de información falsa o ruido [10].

Por otra parte, existe la técnica de Máquinas de Soporte Vectorial (SVM), la cual es una técnica de clasificación supervisada; la teoría detallada de esta metodología se encuentra en la especializada [18], [19], [20]. La principal idea de SVM es separar un par de clases con un hiperplano en particular, el cual maximice la separación entre estas. En la figura 2 se muestra un ejemplo de la función de SVM, en un gráfico bidimensional el hiperplano se convierte en un vector que separa dos clases diferentes, cuya ubicación está establecida por la máxima distancia posibles entre ellas [22].



Fuente: (Brudzewski et al., 2004)

**Figura 1. Gráfico de clasificación típico del SVM. Los puntos encerrados sobre círculos son los soportes del vector**

Diferentes estudios se han llevado mediante el uso de las técnicas de LDA y SVM para la clasificación de productos alimenticios, los cuales son presentados en la Tabla 2. Los modelos de clasificación generalmente son creados en la industria agroalimentaria para discriminar por origen geográfico productos alimenticios, con base en alguna característica medible, que en lo habitual son propiedades fisicoquímicas o sensoriales. En Europa se han llevado a cabo varios estudios con miras a la protección con sellos de calidad DOP de alimentos propios de esa zona, entre estos se destaca el aceite de oliva. Otros estudios en diferentes lugares también han enfocado sus investigaciones a la diferenciación de productos cárnicos, lácteos, mieles o frutas y hortalizas.

**Tabla 2. Reporte bibliográfico sobre métodos de clasificación**

Autor	Estudio	Conclusión
(González Martín et al., 1999)	Clasificación por LDA de aceites de oliva mediante la evaluación del perfil aromático con nariz electrónica.	Los resultados permitieron la creación de un modelo de clasificación de aceites: aceite virgen de oliva, aceite refinado de oliva y aceite de semilla.
(Galdón et al., 2011)	Diferenciación de cultivos de papa basado en su composición química por LDA.	El modelo permitió la diferenciación de los cultivos con base en la medición de 11 parámetros fisicoquímicos.
(Lerma-García et al., 2011)	Diferenciación por LDA de aceite de oliva de acuerdo con su variedad genética mediante la medición del perfil de esteroides por Cromatografía.	El modelo obtenido permitió diferenciar muestras pertenecientes a siete diferentes variedades genéticas españolas con una capacidad de predicción del 97%.



(Papadopoulou et al., 2011)	Evaluación del perfil aromático de carne roja empacada a diferente temperatura mediante nariz electrónica y SVM.	El método proporcionó una correcta clasificación de las carnes de acuerdo con el perfil aromático obtenido a 5 diferentes temperaturas entre 0 y 20°C, la capacidad predicción osciló entre 76-81%.
(Borin et al., 2006)	Cuantificación de adulterantes en leche en polvo, mediante mediciones por espectroscopía de infrarrojo cercano y modelamiento por SVM.	Creación de modelos para la detección de leche en polvo con contenidos de almidón, suero o sacarosa.
(Nashat et al., 2010)	Creación de un sistema inteligente para la inspección de color en biscochos con base en el tiempo de horneado.	Implementación de un sistema de reconocimiento de defectos en color en biscochos con una capacidad de predicción del 85%.
(Alcázar et al., 2012)	Discriminación por origen de cervezas mediante descriptores químicos y SVM.	Creación de un modelo de clasificación de cervezas de tres países Alemania, Portugal y España con una capacidad de del predicción del 99%.
(Zuluaga et al., 2010)	Diferenciación por origen de mieles colombianas por LDA.	Creación de un modelo capaz de diferenciar el origen de mieles provenientes de cuatro de regiones de Colombia.

Fuente: elaboración propia

## Métodos de Regresión

En esta parte de regresión multivariada, se tiene la matriz de datos  $X$  y una matriz de variables cuantitativas  $Y$ , que a diferencia de los métodos de clasificación no expresa una característica o atributo, sino una respuesta numérica, la cual se desea modelar a partir de la información de la matriz  $X$ . Al igual que en los métodos de clasificación, la regresión multivariada requiere las etapas de creación del modelo y validación [2].

Una de las técnicas de regresión más utilizadas en la industria agroalimentaria son los Mínimos Cuadrados Parciales (PLS). Estos encuentran la relación entre la matriz de datos  $X$  con la matriz respuesta  $Y$ . El PLS intenta encontrar factores que cumplan tanto maximizar varianza como lograr correlación [19].

Diferentes estudios se han llevado mediante el uso de las técnicas de PLS para correlacionar la respuesta de una variable dependiente con varias independientes, los cuales se presentan en la Tabla 3. Los métodos de regresión se han empleado ampliamente desde que se comenzaron a utilizar instrumentos analíticos para la descripción de propiedades fisicoquímicas de los alimentos, principalmente humedad, ácidos grasos, carbohidratos y el cálculo de calorías. Los resultados de las técnicas espectrométricas, calorimétricas y cromatográficas, generalmente aprovechan estos métodos para la descripción de tendencias, ya que proporcionan una gran cantidad de información, que sobrepasa las 200 variables. La conclusión obtenida por diferentes investigadores es que la predicción de valores a partir de estas técnicas de regresión reduce los costos operativos, ya que es posible obtener un dato confiable de una propiedad mediante un análisis indirecto, que generalmente es más económico.

**Tabla 3. Reporte bibliográfico sobre métodos de regresión**

Autor	Estudio	Resultado
(Moros et al., 2005)	Determinación del valor energético de bebidas españolas a base de fruta y leche por espectrometría infrarrojo.	Los resultados demostraron que es posible determinar el valor energético a partir de análisis de espectrometría con un error estándar inferior al 2,5%.
(Cozzolino et al., 2005)	Determinación de ácidos grasos libres y el contenido de humedad en aceites de pescado por espectrometría infrarrojo.	Fue posible determinar el contenido de ácidos grasos y humedad a partir de mediciones de espectroscopia. Los coeficientes de regresión oscilaron valores cercanos a 0,95.
(Maggio et al., 2009)	Determinación de ácidos grasos libres en aceite de oliva por espectrometría infrarrojo.	Fue posible determinar el contenido de ácidos grasos y a partir de mediciones de espectroscopia. Los coeficientes de regresión oscilaron valores cercanos a 0,95.
(Mahesar et al., 2010)	Determinación de grasas trans en alimentos basados en cereal Pakistaníes por espectrometría infrarrojo.	Fue posible determinar el contenido de grasas trans a partir de mediciones de espectroscopia. Los coeficientes de regresión oscilaron valores cercanos a 0,99.
(Cerretani et al., 2011)	Determinación de ácidos grasos libres en aceite de oliva por calorimetría diferencial de barrido.	Se encontraron correlaciones aceptables para predecir el contenido de ácido grasos mediante calorimetría, los coeficientes de regresión oscilaron entre 0.77 y 0.94 para los ácidos palmítico, esteárico, oleico y linoleico.

Fuente: elaboración propia

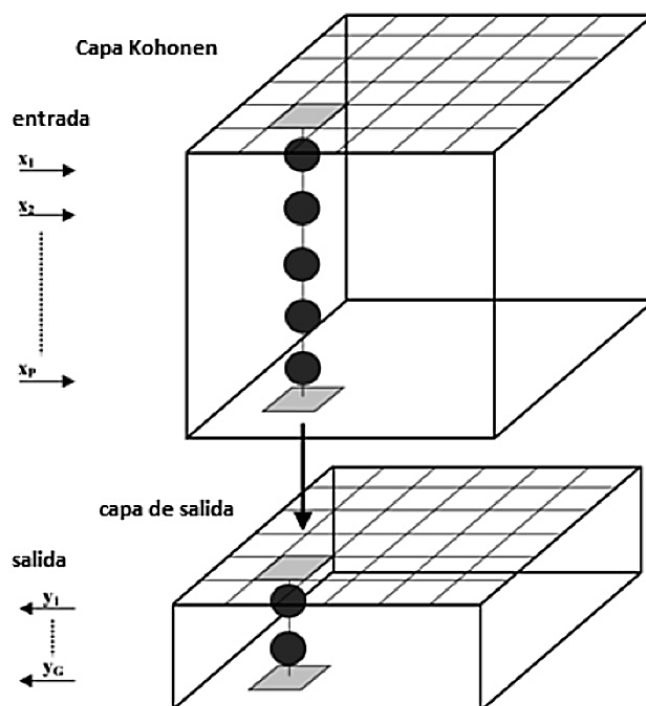
### **Métodos de Aprendizaje y Optimización**

Estos métodos son los más modernos para el análisis de información de sistemas complejos, su utilidad se basa en la capacidad de “aprender” y reorganizar la información de manera tal que siempre se conserve aquella relevante y se elimine rápidamente la que no aporta mucho para el sistema. Estas técnicas, muy útiles para modelar fenómenos también, han sido aprovechadas para optimizar matrices con un alto grado de variables, mediante la eliminación *a priori* de aquellas que no dan información relevante o aportan únicamente ruido a la matriz. Dentro de esas técnicas, las redes neuronales artificiales y los algoritmos genéticos son los más reconocidos.

Las redes neuronales por contrapropagación (CP-ANN) son un método inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida, consistente de dos capas, una capa llamada Kohonen, la cual puede ser usada para el análisis estructural de los datos y una capa de salida. Esta técnica está basada en la búsqueda de similitudes en las muestras y permite proyectarlas en un espacio topológico donde las muestras similares están cerca las unas de las otras y las disímiles aparte (Ver Figura 3) [5], [7], [8].

Las redes neuronales artificiales por contrapropagación pueden ser consideradas un método de aprendizaje de clases, ya que puede reconocer muestras que no pertenecen a ninguna de las clases. Esto ocurre cuando las muestras son asignadas a neuronas donde los pesos de la capa de salida son similares, es decir, la neurona no puede ser asignada a una clase específica. Las variables originales son usualmente pretratadas escalándolas en un rango entre 0 y 1 para hacerlas comparables con el peso de las redes [7].





Fuente: (Ballabio et al., 2009)

Figura 3. Representación gráfica de las capas de entrada y salida de la red neuronal

Por otra parte, los algoritmos genéticos (GA) son descritos ampliamente por Leardi y cols. El GA básicamente inicia con una población aleatoria inicial de cromosomas. Cada cromosoma está compuesto de un conjunto binario de genes, los cuales representan los parámetros a ser optimizados maximizando una determinada función de ajuste (o criterio de optimización). Posteriormente, un proceso de evolución es simulado y se obtienen nuevas cromosomas (descendientes) a partir del apareamiento de los cromosomas de la población inicial mediante operadores genéticos (entrecruzamiento y mutación).

Los peores cromosomas en la población serán reemplazados con los nuevos, si sus valores de ajuste son peores que los de los nuevos. El proceso evolutivo es repetido por un número definido de evaluaciones y solamente los cromosomas con los valores más altos de ajuste permanecerán en la población final [23]. Es un método de optimización, ya que así como lo hace la selección natural, de acuerdo con algún criterio se decide cuáles son los individuos (en este caso variables) más adaptados, que sobrevivirán, y cuáles los menos aptos que serán descartados.

Diferentes estudios se han llevado mediante el uso de las técnicas de ANN y GA para modelar y optimizar sistemas complejos de datos, los cuales son presentados en la Tabla 4. En el caso de modelamiento por redes neuronales, los estudios se han enfocado hacia investigaciones más complejas, que las técnicas anteriormente mencionadas no son capaces de describir adecuadamente. Su uso se ha visto aplicado a la descripción de cinéticas de crecimiento microbiano, cinética enzimática, procesos de transferencia de masa y calor, entre otros. Por otra parte, los algoritmos genéticos se han empleado hacia la optimización de etapas de transformación, con el fin de encontrar los valores de las variables que no solamente mejoren el proceso, sino que también minimicen el costo operativo. En ese orden de ideas, esta técnica se utiliza generalmente en el mejoramiento de procesos de secado, absorción, destilación, hidrogenación, entre otros. Cabe aclarar que estas técnicas también son ajustables en la caracterización de alimentos mediante la medición de propiedades fisicoquímicas, sensoriales o de vida útil.

**Tabla 4. Reporte bibliográfico sobre métodos de aprendizaje y optimización**

Autor	Estudio	Resultado
(Ni et al., 1999)	Creación de un modelo por ANN para describir la cinética de degradación de antioxidantes.	Fue posible encontrar un modelo que expresara la tasa de degradación para Hidroxianisol butilado (BHA) hidroxitolueno butilado (BHT) y propil galato (PG).
(Keeratipibul et al., 2011)	Predicción de la presencia de coliformes en frutas y lechuga por ANN, después de la sanitización con varios agentes.	Se logró la mejor discriminación de vegetales respecto con los agentes de sanitización. Se encontró que el mejor agente es el ácido peracético.
(Geeraerd et al., 1998)	Modelamiento por ANN del crecimiento bacteriano en productos congelados.	Es posible crear un modelo de predicción del crecimiento de microorganismos. Los factores que más influenciaron el modelo son pH y %NaCl.
(Gonçalves et al., 2005)	Modelamiento del proceso de esterilización de enlatados por ANN.	Se creó un modelo para monitorear el proceso tomando como datos iniciales el tiempo de proceso y las temperaturas de los puntos fríos. La respuesta consistía en la temperatura del punto frío a un tiempo determinado.
(Torrecilla et al., 2005)	Creación de modelos por ANN para la optimización de procesos en altas presiones en alimentos.	Se creó un modelo para predecir la temperatura máxima y mínima alcanzada en un alimento después de la presurización y el tiempo necesario para alcanzar el equilibrio térmico.
(Razmi-Rad et al., 2007)	Predicción de las propiedades reológicas de la masa de pan a partir de la composición química de la harina de trigo por ANN.	Fue posible crear un modelo de predicción de las propiedades de la masa a partir de la evaluación de los parámetros contenido de proteína, gluten, valor de sedimentación y número de caída.
(Llave et al., 2012)	Predicción por ANN de la temperatura del punto más frío en la esterilización de alimentos.	Se creó un modelo de predicción de la temperatura del punto más frío en función del tiempo de esterilización.
(Penza et al., 2004)	Creación de un modelo de clasificación por ANN de vinos italianos mediante análisis por nariz electrónica.	Se diferenciaron nueve diferentes tipos de vino de acuerdo a su origen geográfico, con una capacidad de predicción del modelo del 78%.
(Cimpoiou et al., 2011)	Predicción de la actividad antioxidante de té usando ANN.	Se creó un modelo de predicción de la capacidad antioxidante a partir de los datos de contenido de flavonoides, catequinas y metil-xantinas con un error inferior al 4%.
(Cevoli et al., 2011)	Clasificación de quesos Pecorino por ANN usando una nariz electrónica y cromatografía de gases.	El modelo creado logró predecir el tiempo de almacenamiento de los quesos mediante las características de aroma descritas.
(Lou et al., 2001)	Predicción de la inactivación térmica de bacterias usando ANN.	Se creó un modelo que evalúa la cinética de destrucción de microorganismos a partir de las variables: temperatura, pH y actividad de agua.

(Continuación de la tabla)

(Yuzgec et al., 2006)	Control predictivo de un proceso de secado usando GA.	Se creó un modelo de optimización para la predicción del perfil de secado de levadura de panadería. Los resultados mostraron un incremento en la calidad y una reducción en el consumo de energía y tiempo de secado.
(Izadifar et al., 2007)	Optimización del proceso de hidrogenación de aceite vegetal usando GA.	Se creó un modelo de optimización que encuentra las condiciones de proceso que minimice la formación de isómeros trans y maximice la producción de ácido cis-oleico.
(González-Sáiz et al., 2008)	Modelamiento de la transferencia de masa Gas-Líquido / Líquido-Gas en la producción de vinagre.	Se creó un modelo por GA con los cuales fue posible encontrar los parámetros óptimos para el proceso de fermentación para la elaboración de vinagre, a partir de los cuales se obtuvieron los coeficientes de transferencia de masa.
(Buratti et al., 2007)	Predicción de descriptores sensoriales instrumentales de vinos rojos italianos.	Se creó un modelo a partir del cual se establecieron los parámetros de los instrumentos nariz electrónica, lengua electrónica y espectrofotómetro que permiten predecir los atributos de los vinos de acuerdo con su origen.
(Jaya Shankar et al., 2004)	Optimización de proceso de extrusión mediante GA.	Se creó un modelo que logró encontrar los parámetros óptimos para la extrusión de arroz. Las variables modeladas fueron relación de expansión, densidad, dureza y solubilidad en agua.

## Conclusiones

El creciente número de artículos científicos en los cuales las diferentes técnicas estadísticas multivariadas son aplicadas en diversos campos de la industria agroalimentaria muestra la efectividad y validez de estos métodos. La ventaja de la estadística multivariada sobre las técnicas “clásicas” es que reduce en muchos casos la complejidad del problema. Los resultados abarcan desde la simple exploración de información hasta la creación de modelos de optimización, se destaca también la posibilidad de usar métodos de regresión y clasificación para la descripción de fenómenos. Sin embargo, debe resaltarse que no siempre es posible definir un modelo, encontrar un óptimo o simplemente definir un método de exploración de información simple que sea útil en todas las aplicaciones, por lo cual depende de la experticia del investigador la selección de la técnica para obtener resultados confiables.

## Referencias Bibliográficas

- [1] Kvalheim o. (1987). *Latent-structure decompositions (projections) of multivariate data*. Chemometrics and Intelligent Laboratory Systems Vol. 2, pp.283-290. Leardi R. (2001). *Genetic algorithms in chemometrics and chemistry: a review*. Journal of Chemometrics Vol. 15, No.7 pp.559-569.
- [2] Stahle L. y Wold S. (1987). *Partial least squares analysis with cross-validation for the two-class problem: a monte carlo study* Journal of chemometrics Vol. 1, pp.185-196.
- [3] Wold S., Esbesen K. y Geladi P.(1987). *Principal Component Analysis*. Chemometrics and Intelligent-Laboratory Systems Vol. 2, pp.3752.

- [4] McLachlan G. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. NY, USA, Wiley.
- [5] Zupan J. (1994). *Introduction to artificial neural network (ann) methods: what they are and how to use them*. Acta Chimica Slovenica Vol. 41, pp.327-352.
- [6] Poulsen J. y French A. (1996). *Discriminant Function Analysis (DA)*. San Francisco State University, USA Vol. Razmi-Rad E., Ghanbarzadeh B., Mousavi S. M., Emam-Djomeh Z. y Khazaei J. (2007).
- [7] Zupan J., Novic M. y Ruisanchez I. (1997). *Kohonen and counterpropagation artificial neural networks in analytical chemistry*. Chemometrics and Intelligent Laboratory Systems Vol. 38, pp.1-23.
- [8] Zupan J. y Gasteiger J. (1999). *Neural networks in chemistry and drug design*. Weinheim, Germany, VCH.
- [9] McLachlan G. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. New Jersey, USA., John Wiley & Sons, Inc. Moros J., Iñón F.A., Garrigues S. y de la Guardia M. (2005). *Determination of the energetic value of fruit and milk-based beverages through partial-least-squares attenuated total reflectance-Fourier transform infrared spectrometry*. Analytica Chimica Acta Vol. 538, No.1-2, pp.181-193.
- [10] Bereton R. (2007). *Applied Chemometrics for Scientists*. USA, John Wiley & Sons.
- [11] Garcia D. (2009). *Fault detection using Principal Component Analysis (PCA) in a wastewater treatment plant (WWTP)*. Valladolid, Spain, Department of Systems Engineering and Automatic Control. University of Valladolid.
- [12] Massart D. L., Vandeginste B., Buydens L., De Jong S., Lewi P. y J. S. (1997). *Handbook of Chemometrics and Qualimetrics: Part A*. Amsterdam, The Netherlands, Elsevier.
- [13] Ballabio D. (2006). *Chemometric characterisation of physical-chemical fingerprints of food products*. Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche. Milano, Università degli Studi di Mila- no. Dottorato di Ricerca in Biotecnologie degli Alimenti.
- [14] WOK (2012). *Web of knowledge*. Recuperado de [http://apps.webofknowledge.com/WOS\\_General-Search\\_input.do?product=WOS&search\\_mode=GeneralSearch&SID=3C1o@fn8K4O325jaJ3o&preferen cesSaved=&highlighted\\_tab=WOS](http://apps.webofknowledge.com/WOS_General-Search_input.do?product=WOS&search_mode=GeneralSearch&SID=3C1o@fn8K4O325jaJ3o&preferen cesSaved=&highlighted_tab=WOS).
- [15] SIC (2004). *Denominaciones de Origen en la Comunidad Andina*. Bogotá, Colombia., Superintendencia de Industria y Comercio.
- [16] James M. (1985). *Classification Algorithms*. London, UK, Collins.
- [17] Hand D. (1997). *Construction and Assessment of Classification Rules*. Chichester, UK, Wiley.
- [18] Duda R., Hart P. y Stork D. (2000). *Pattern Classification and Scene Analysis*. Wiley, 2nd Ed. Vol.
- [19] Frank I. y Friedman J. (1989). *Classification: oldtimers and newcomers*. Journal of Chemometrics Vol. 3, pp.463-475.
- [20] Webb A. (2002). *Statistical Pattern Recognition*. Arnold, 2nd Ed. Vol.
- [21] Distanto C., Ancona N. y Siciliano P. (2003). *Support Vector Machines for Olfactory Signals Recognition*. Sensors and Actuators B Vol. 88, pp.30-39.
- [22] Brudzewski K., Osowski S. y Markiewicz T. (2004). *Classification of milk by means of an electronic nose and SVM neural network*. Sensors and Actuators B: Chemical Vol. 98, No.2-3, pp.291-298.
- [23] Leardi R. y Lupianez-Gonzalez (1998). *Genetic algorithms applied to feature selection in PLS regression: how and when to use them*. Chemometrics and Intelligent Laboratory Systems Vol. 41, No.2, pp.195-207
- [24] Alcázar Á., Jurado J. M., Palacios-Morillo A., de Pablos F. y Martín M. J. (2012). *Recognition of the geographical origin of beer based on support vector machines applied to chemical descriptors*. Food Control Vol. 23, No.1, pp.258-262.

- [25] Baardseth P., Nilsen T. y Vogt G. (1995). *Roll-in shortenings effects on Danish pastries sensory properties studied by principal component analysis*. LWT - Food Science and Technology Vol. 28, No.1, pp.72-77
- [26] Ballabio D. y Todeschini R. (2009). *Chapter 4 - Multivariate Classification for Qualitative Analysis*. Infrared Spectroscopy for Food Quality Analysis and Control. San Diego, Academic Press: 83-10.
- [27] Borin A., Ferrão M. F., Mello C., Maretto D. A. y Poppi R. J. (2006). *Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk*. Analytica Chimica Acta Vol. 579, No.1, pp.25-32.
- [28] Buratti S., Ballabio D., Benedetti S. y Cosio M. S. (2007). *Prediction of Italian red wine sensorial descriptors from electronic nose, electronic tongue and spectrophotometric measurements by means of Genetic Algorithm regression models*. Food Chemistry Vol. 100, No.1, pp.211-218.
- [29] Cerretani L., Maggio R. M., Barnaba C., Toschi T. G. y Chiavaro E. (2011). *Application of partial least square regression to differential scanning calorimetry data for fatty acid quantitation in olive oil*. Food Chemistry Vol. 127 No.4, pp.1899-1904.
- [30] Cevoli C., Cerretani L., Gori A., Caboni M. F., Gallina Toschi T. y Fabbri A. (2011). *Classification of Pecorino cheeses using electronic nose combined with artificial neural network and comparison with GC-MS analysis of volatile compounds*. Food Chemistry Vol. 129, No.3, pp.1315-1319.
- [31] Cimpoi C., Cristea V.-M., Hosu A., Sandru M. y Seserman L. (2011). *Antioxidant activity prediction and classification of some teas using artificial neural networks*. Food Chemistry Vol. 127 No.3, pp.1323-1328.
- [32] Cozzolino D., Murray I., Chree A. y Scaife J. R. (2005). *Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy*. LWT - Food Science and Technology Vol. 38, No.8, pp.821-828.
- [33] De Belle N., De Smedt V. y De Baerdemaeker J. (2000). *Principal component analysis of chewing sounds to detect differences in apple crispness*. Postharvest Biology and Technology Vol. 18, No.2, pp.109-119.
- [34] E. M. y Romero C. D. (2011). *Differentiation of potato cultivars experimentally cultivated based on their chemical composition and by applying linear discriminant analysis*. Food Chemistry In press.
- [35] Geeraerd A. H., Herremans C. H., Cenens C. y Van Impe J. F. (1998). *Application of artificial neural networks as a non-linear modular modeling technique to describe bacterial growth in chilled food products*. International Journal of Food Microbiology Vol. 44, No.1-2, pp.49-68.
- [36] Gonçalves E. C., Minim L.A., Coimbra J. S. R. y Minim V.P.R. (2005). *Modeling sterilization process of canned foods using artificial neural networks*. Chemical Engineering and Processing: Process Intensification Vol. 44, No.12, pp.1269-1276.
- [37] González-Sáiz J.-M., Pizarro C. y Garrido-Vidal D. (2008). *Modelling gas-liquid and liquid-gas transfers in vinegar production by genetic algorithms*. Journal of Food Engineering Vol. 87 No.1, pp.136-147
- [38] González Martín Y., Luis Pérez Pavón J., Moreno Cordero B. & García Pinto C. (1999). *Classification of vegetable oils by linear discriminant analysis of Electronic Nose data*. Analytica Chimica Acta Vol. 384, No.1, pp.83-94.
- [39] Guerreiro J. S., Barros M., Fernandes P., Pires P. y Bardsley R. (2012). *Principal component analysis of proteolytic profiles as markers of authenticity of PDO cheeses*. Food Chemistry Vol., No.0.
- [40] Díez L. M. (2011). *Principal component analysis (PCA) and multiple linear regression (MLR) statistical tools to evaluate the effect of E-beam irradiation on ready-to-eat food*. Journal of Food Composition and Analysis Vol. 24, No.3, pp.456-464.
- [41] Horimoto Y. y Nakai S. (1998). *Classification of pasteurized milk using principal component similarity analysis of off-flavours*. Food Research International Vol. 31, No.4, pp.279-287



- [42] Izadifar M. y Jahromi M. Z. (2007). *Application of genetic algorithm for optimization of vegetable oil hydrogenation process*. Journal of Food Engineering Vol. 78, No.1, pp.1- 8.
- [43] Jaya Shankar T. y Bandyopadhyay S. (2004). *Optimization of Extrusion Process Variables Using a Genetic Algorithm*. Food and Bioproducts Processing Vol. 82, No.2, pp.143-150.
- [44] Kallithraka S., Arvanitoyannis I. S., Kefalas P, El-Zajouli A., Soufleros E. y Psarra E. 2001). *Instrumental and sensory analysis of Greek wines; implementation of principal component analysis (PCA) for classification according to geographical origin*. Food Chemistry Vol. 73, No.4, pp.501-514.
- [45] Keeratipibul S., Phewpan A. y Lursinsap C. (2011). *Prediction of coliforms and Escherichia coli on tomato fruits and lettuce leaves after sanitizing by using Artificial Neural Networks*. LWT - Food Science and Technology Vol. 44, No.1, pp.130-138.
- [46] Lerma-García M. J., Simó-Alfonso E. F., Méndez A., Lliberia J. L. y Herrero-Martínez J. M. (2011). *Classification of extra virgin olive oils according to their genetic variety using linear discriminant analysis of sterol profiles established by ultra-performance liquid chromatography with mass spectrometry detection*. Food Research International Vol. 44, No.1, pp.103 -108.
- [47] Lou W. y Nakai S. (2001). *Application of artificial neural networks for predicting the thermal inactivation of bacteria: a combined effect of temperature, pH and water activity*. Food Research International Vol. 34, No.7 pp.573-579.
- [48] Llave Y.A., Hagiwara T. y Sakiyama T. (2012). *Artificial neural network model for prediction of cold spot temperature in retort sterilization of starch-based foods*. Journal of Food Engineering Vol. 109, No.3, pp.553-560.
- [49] Maggio R. M., Kaufman T. S., Carlo M. D., Cerretani L., Bendini A., Cichelli A. y Compagnone D. (2009). *Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares*. Food Chemistry Vol. 114, No.4, pp.1549-1554.
- [50] Mahesar S. A., Kandhro A. A., Cerretani L., Bendini A., Sherazi S. T. H. y Bhangar M. I. (2010). *Determination of total trans fat content in Pakistani cereal-based foods by SB-HATR FTIR spectroscopy coupled with partial least square regression*. Food Chemistry Vol. 123, No.4, pp.1289-1293.
- [51] Nashat S. y Abdullah M. Z. (2010). *Multi-class colour inspection of baked foods featuring support vector machine and Wilk's analysis*. Journal of Food Engineering Vol. 101, No.4, pp.370-380.
- [52] Ni Y. y Liu C. (1999). *Artificial neural networks and multivariate calibration for spectrophotometric differential kinetic determinations of food antioxidants*. Analytica Chimica Acta Vol. 396, No.2-3, pp.221-230.
- [53] Papadopoulou O. S., Tassou C. C., Schiavo L., Nychas G.-J. E. y Panagou E. Z. (2011). *Rapid Assessment of Meat Quality by Means of an Electronic Nose and Support Vector Machines*. Procedia Food Science Vol. 1, No.0, pp.2003-2006.
- [54] Pardo M. y Sberveglieri G. (2005). *Classification of electronic nose data with support vector machines*. Sensors and Actuators B Vol. 107 pp.730-737
- [55] Patras A., Brunton N. P., Downey G., Rawson A., Warriner K. y Gernigon G. (2011). *Application of principal component and hierarchical cluster analysis to classify fruits and vegetables commonly consumed in Ireland based on in vitro antioxidant activity*. Journal of Food Composition and Analysis Vol. 24, No.2, pp.250-256.
- [56] Penza M. y Cassano G. (2004). *Chemometric characterization of Italian wines by thin-film multisensors array and artificial neural networks*. Food Chemistry Vol. 86, No.2, pp.283-296.
- [57] *Prediction of rheological properties of Iranian bread dough from chemical composition of wheat flour by using artificial neural networks*. Journal of Food Engineering Vol. 81, No.4, pp.728-734.
- [58] Rodríguez-Campos J., Escalona-Buendía H. B., Orozco-Avila I., Lugo-Cervantes E. y Jaramillo-Flores M. E. (2011). *Dynamics of volatile and non-volatile compounds in cocoa (Theobroma cacao L.) during fermentation and drying processes using principal components analysis*. Food Research International Vol. 44, No.1, pp.250-258.



- [60] Rodríguez-Delgado M.-Á., González-Hernández G., Conde-González J.-E. a. y Pérez- Trujillo J.-P. (2002). *Principal component analysis of the polyphenol content in young red wines*. Food Chemistry Vol. 78, No.4, pp.523-532.
- [61] Shin E.-C., Craft B. D., Pegg R. B., Phillips R. D. y Eitenmiller R. R. (2010). *Chemometric approach to fatty acid profiles in Runner-type peanut cultivars by principal component analysis (PCA)*. Food Chemistry Vol. 119, No.3, pp.1262-1270.
- [62] Torrecilla J. S., Otero L. & Sanz P.D. (2005). *Artificial neural networks: a promising tool to design and optimize high-pressure food processes*. Journal of Food Engineering Vol. 69, No.3, pp.299-306.
- [63] Vandeginste B., Massart D. L., Buydens L., De Jong S., Lewi P. y J. S. (1998). *Handbook of Chemometrics and Qualimetrics. Part B*. Amsterdam, The Netherlands, Elsevier.
- [64] Yuzgec U., Becerikli Y. y Türker M. (2006). *Nonlinear predictive control of a drying process using genetic algorithms*. ISA Transactions Vol. 45, No.4, pp.589-602.
- [65] Zuluaga C., Diaz C., Henao N. & Quicazan M. (2010). *Diferenciación por origen de mieles colombianas de acuerdo a su contenido mineral y perfil aromático*.
- [66] Encuentro Nacional de Investigación y Desarrollo. Universidad Nacional de Colombia.