



# **UNIVERSIDAD NACIONAL DEL SUR**

**TESIS MAGÍSTER EN CIENCIA Y TECNOLOGÍA DE LOS ALIMENTOS**

***DETERMINACIÓN DE LA ADULTERACIÓN DE ALIMENTOS MEDIANTE LA  
IMPLEMENTACIÓN DE METODOLOGÍAS ANALÍTICAS BASADAS EN  
TÉCNICAS ESPECTROSCÓPICAS Y ANÁLISIS MULTIVARIANTE***

**GIMENA JÉSSICA TRENTANNI HANSEN**

**BAHÍA BLANCA**

**ARGENTINA**

**2018**

## PREFACIO

Esta tesis es presentada como parte de los requisitos para optar al grado académico de Magister en Ciencia y Tecnología de los Alimentos de la Universidad Nacional del Sur y no ha sido presentada para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en el Instituto de Química del Sur INQUISUR-UNS-CONICET Departamento de Química de la Universidad Nacional del Sur, durante el periodo comprendido entre el 2 de Abril del 2015 y el 19 de Febrero del 2018, bajo la dirección de la Dra. Liliana Albertengo, Profesora Asociada, y la codirección de la Dra. Carolina Di Anibal, docente UNS e Investigadora Adjunta CONICET, ambas del Área III, Química Analítica, del Departamento de Química de la Universidad Nacional del Sur.

Bqca. Gimena Jéssica Trentanni Hansen



UNIVERSIDAD NACIONAL DEL SUR  
Secretaría General de Posgrado y Educación Continua

La presente tesis ha sido aprobada el ..../..../..... , mereciendo la calificación de .....(.....)

*Le dedico esta tesis al gran  
amor de mi vida, mi hija*

*Agostina*

## **AGRADECIMIENTOS**

Realizar una tesis de Magister implica recorrer un determinado camino, que si bien es personal, involucra el aporte de un gran número de personas. Por ello, es para mí un verdadero placer utilizar este espacio para expresarles mi más sincero agradecimiento:

A la Universidad Nacional del Sur, por brindarme las instalaciones para poder desarrollarla.

A la Dra. Liliana Albertengo por aceptar la dirección y a la Dra. Carolina Di Anibal por su codirección de la presente tesis, por su apoyo y confianza en mí, su capacidad para guiar mi trabajo, sus constantes aportes académicos, su generosidad y por compartir continuamente su sabiduría. Mi agradecimiento no sólo a nivel profesional, sino también a nivel personal.

Al grupo de Bromatología por su apoyo.

Al Dr. Paulo Henrique Gonçalves Dias Diniz de la Universidad Estatal de Paraíba, Brasil, por su aporte brindado en la parte quimiométrica.

En el plano personal quiero agradecer muy especialmente a la persona que le da sentido pleno a mi vida, quien convierte mis problemas en suyos y me hace completamente feliz. Gracias Nicolás por hacer de mí todos los días una mejor persona. En este logro, como en cualquier otro, ocupas un lugar primordial.

Quiero agradecer muy especialmente a mis padres por incentivar me siempre, darme su apoyo absoluto y brindarme sencillamente todo para hoy ser quien soy. Mi agradecimiento es eterno hacia ustedes, y no cabe en ningún texto posible. Simplemente gracias. A mi hermano por siempre estar presente cuando lo necesito.

## RESUMEN

Esta tesis está focalizada en el desarrollo de métodos analíticos de screening para determinar la adulteración de especias culinarias, como pimentones, con colorantes Sudan I, II, III, IV y Para-Red. La utilización de estos colorantes en alimentos está prohibida ya que son potencialmente carcinógenos. Los métodos utilizados habitualmente para detectar estos colorantes en matrices alimentarias requieren tiempos largos de análisis e instrumentación costosa. En esta tesis se propone desarrollar un método de *screening* como alternativa útil, con beneficios como la reducción del tiempo y costo de análisis, siendo de especial utilidad en situaciones que requieren una rápida respuesta, por la peligrosidad asociada a esta adulteración.

La metodología analítica desarrollada está basada en el uso de técnicas espectroscópicas y tratamiento multivariante de los datos obtenidos. Para ello se evaluó la espectroscopía de Infrarrojo cercano (NIR) junto con análisis multivariante, para establecer modelos de clasificación. Con el fin de obtener una idea previa del agrupamiento natural de los datos se implementó un análisis exploratorio, para luego construir una regla de clasificación que permita determinar si una muestra está o no adulterada, y en caso afirmativo, predecir qué tipo de adulterante contiene. Posteriormente se utilizaron herramientas quimiométricas con el objetivo de mejorar los resultados de clasificación como selección de variables, evaluando dos métodos de diferente naturaleza.

La metodología propuesta representa una herramienta de screening fiable, rápida, simple y asequible para determinar adulteraciones alimentarias de gran impacto en la sociedad como lo es la ingesta de colorantes de reconocida peligrosidad. Cabe resaltar que la misma podría ser implementada en laboratorios de control de rutina de alimentos, tanto gubernamentales como privados. En este sentido, se está trabajando en la elaboración de un dispositivo portátil integrado para análisis in-situ, con el fin de poder detectar los colorantes de la familia Sudan de forma rápida y económica en los puntos de interés.

## **ABSTRACT**

This thesis is focused on the development of analytical screening methods for determining the adulteration of culinary spices such as paprika with Sudan I to IV and Para-Red dyes. These dyes are forbidden to be used in foods because they are potential carcinogens. The methods commonly used to detect these dyes in foods are time-consuming and expensive. Therefore, a screening methodology is proposed as alternative tool with the advantage of reducing both time and cost analysis, being useful in situations that require a rapid response.

The proposed methodology is based on spectroscopic techniques and multivariate analysis, specifically, Near Infrared (NIR) spectroscopy with multivariate classification. An exploratory analysis was implemented with the aim to have an initial view of the samples innate clustering. Then a classification rule was built to predict if an unknown sample is or not adulterated, and if so, predict which type of adulterant it contains. Finally, in order to improve the obtained classification results, different variable selection strategies were implemented.

The developed analytical methodology is a rapid, simple and reliable screening tool for determining food adulteration that have high impact in public health derived from the intake of dangerous dyes. It must be highlighted that it could be implemented in food control routine laboratories and government laboratories. In this context, it is being developed an integrated-portable device for in-situ analysis, with the aim to detect Sudan dyes in a rapid and low-cost way.

# ÍNDICE

<b>1. INTRODUCCIÓN</b>	
1.1 Alimentos y adulteraciones	1
1.2 Especies culinarias	3
1.2.1 Pimentón	4
1.3 Colorantes Sudan y Para-Red	8
1.3.1 Detección colorantes Sudan y Para-Red	11
1.4 Métodos de screening	13
1.5 Espectroscopia de Infrarrojo Cercano NIR	14
1.6 Análisis multivariante	18
1.6.1 Pre-procesamiento de espectros	19
1.6.2 Técnicas de reconocimiento de patrones	22
1.6.2.1 Técnicas de reconocimiento de patrones no supervisadas	22
1.6.2.2 Técnicas de reconocimiento de patrones supervisadas	23
1.6.3 Validación cruzada	25
1.6.4 Selección de variables	26
1.6.5 Evaluación de la metodología desarrollada	27
<b>2. OBJETIVOS</b>	
2.1 Objetivos	31
<b>3. MATERIALES Y METODOS</b>	
3.1 Muestras y adulterantes	32
3.2 Preparación de muestras	32
3.3 Instrumentación y software	32
3.4 Análisis multivariante	35

<b>4. RESULTADOS Y DISCUSION</b>	
4.1 Tratamiento de muestras	36
4.2 Obtención de espectros NIR	40
4.3 Pre-procesamiento de espectros	43
4.4 Modelos de clasificación multivariante	46
4.5 Selección de variables i-modelos	49
4.6 Evaluación del modelo de clasificación obtenido con i-PCA	51
4.7 Evaluación de la capacidad predictiva del modelo	55
4.8 Selección de variables SPA-LDA	56
<b>5. CONCLUSIONES</b>	
5.1 Conclusiones	59
<b>6. BIBLIOGRAFIA</b>	
6.1 Bibliografía	61

# 1. INTRODUCCIÓN

## **1.1 Alimentos y adulteraciones**

Los alimentos son productos naturales o elaborados que, debido a la calidad tanto nutricional como de inocuidad y de aceptación sensorial, contribuyen al mantenimiento de las funciones normales del organismo, es decir, al mantenimiento de la salud. Desde el punto de vista químico, los alimentos son una matriz compleja y heterogénea formada por diversos componentes. La evaluación de la calidad, características sensoriales, valor nutricional y estabilidad de productos alimentarios son el principal interés de los químicos analíticos que trabajan, tanto en industrias de procesamiento de alimentos como en laboratorios de control de calidad gubernamentales o privados. El rol principal del análisis de alimentos está enfocado en asegurar una correcta calidad y característica higiénica de los mismos. El desafío constante comprende la implementación de normas reguladoras y legislación actualizada, asociadas a la utilización de metodologías analíticas, para determinar modificaciones y adulteraciones de forma rápida y que sean adaptables a laboratorios de rutina.

La cadena de suministros de alimentos se ha ido transformando progresivamente, siendo cada vez más global y compleja, surgiendo nuevos y desafiantes riesgos en materia de seguridad alimentaria. Existen una gran variedad de peligros asociados a los alimentos, y tanto los ingredientes como los aditivos presentan una compleja problemática porque, generalmente, no tienen una característica funcional única o que pueda evaluarse visualmente, lo que no permite su discriminación de otros ingredientes similares o de posibles adulterantes (Cozzolino, 2015).

El uso de aditivos alimentarios exige el cumplimiento, de acuerdo con la norma general del Codex (CODEX STAN 192-1995), de los siguientes principios: inocuidad, justificación de uso, buenas prácticas de manufactura (BMP) y especificaciones de identidad y pureza. La utilización de aditivos no autorizados para consumo humano conlleva a la contaminación y adulteración de los alimentos. Como está considerado en la definición del Código Alimentario Argentino (CAA) en el Capítulo 1 en el artículo 6, un alimento adulterado es aquel que ha sido privado, en forma parcial o total, de sus elementos útiles o característicos, reemplazándolos o no por otros inertes o extraños; que ha sido adicionado de aditivos no autorizados o sometidos a tratamientos de cualquier naturaleza para disimular u ocultar alteraciones, deficiente calidad de materias primas o defectos de elaboración. Las adulteraciones alimentarias no constituyen un problema contemporáneo, son tan antiguas como los sistemas de producción y procesamiento de alimentos, con la diferencia que se han transformado en más

sofisticadas, siguiendo o anticipándose a los avances en el campo de las ciencias analíticas. Ejemplos de adulteraciones alimentarias de gran impacto mundial incluyen, en Inglaterra (2005) la presencia de colorantes no autorizados en chile en polvo que fueron utilizados en salsas Worcestershire, en China (2007) el aumento del contenido proteico en leche con el agregado de gluten de trigo y melamina lo cual derivó en muertes infantiles y miles de hospitalizaciones, en India (2012) leche adulterada con detergente, grasa y urea. Otras adulteraciones de alto impacto en la salud pública fueron la presencia de óxido de plomo ( $Pb_3O_4$ ) en pimentones y pimienta de cayena y la contaminación de carne de cerdo con dioxinas (Irlanda, 2012) que constituyó la mayor crisis que afectó la economía y la salud nacional. Recientemente, en Europa, se impidió la importación de miel asiática por la presencia ilegal de antibióticos y también de mieles fabricadas a partir de edulcorantes artificiales, y en 2013 en Taiwán, se detectaron contaminaciones de aceite de oliva con aceites baratos de diferentes orígenes y clorofilina, así como de bebidas alcohólicas, leche y arroz (Lohumi et al., 2015).

La figura 1.1 muestra el porcentaje de adulteraciones reportadas en diferentes tipos de alimentos en la literatura científica. Se puede observar que el mayor porcentaje corresponde a aceites, seguido de leche y en tercer lugar, con valores muy similares, productos derivados de frutas (12%) y especias (11%) (Moore et al, 2012).

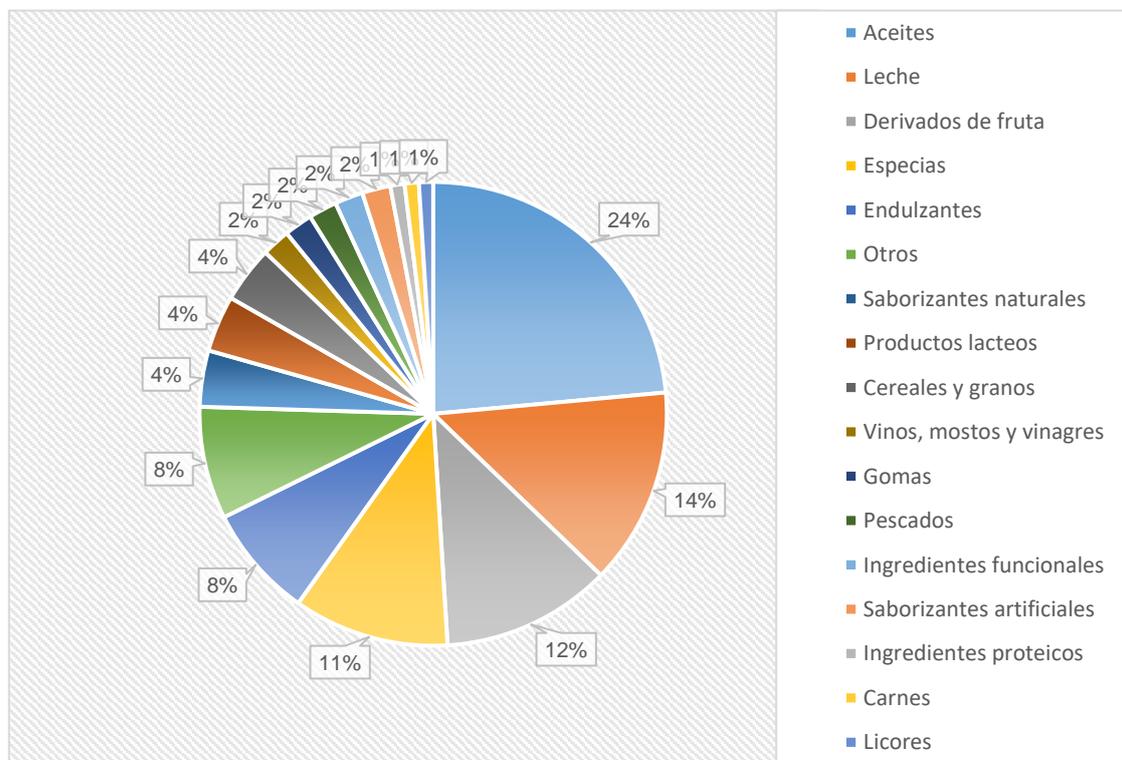


Figura 1.1. Adulteraciones en diferentes grupos de alimentos.

Entre las materias primas sujetas a un mayor número de adulteraciones se encuentran especias, aceites comestibles, miel, leche y productos lácteos, frutas y jugos de fruta, café, harina, productos cárnicos.

El peligro de adulteraciones a lo largo de la cadena de suministro de alimentos es relevante, ya que los consumidores no pueden reducirlo sustancialmente durante la preparación de alimentos. Por consiguiente, en materia de seguridad alimentaria se requiere de una vigilancia continua, siendo de primordial importancia la investigación y desarrollo de técnicas analíticas rápidas y simples para detectar adulteraciones y contaminaciones en alimentos y en materias primas utilizadas en su elaboración. En este contexto, los métodos de screening rápidos, eficientes y que puedan ser adaptados a dispositivos manuales, representarían una herramienta muy útil para ser utilizada en puntos de vulnerabilidad a lo largo de la compleja red de suministros de materias primas y alimentos (Ellis et al., 2015).

## **1.2 Especias culinarias**

Desde tiempos remotos, las especias y hierbas aromáticas han sido utilizadas como aditivos en preparaciones culinarias, medicinales y cosméticas. Las especias culinarias están presentes en la mayoría de los alimentos procesados, incluyendo productos listos para consumir, y además son usualmente utilizadas por los consumidores para aumentar la calidad organoléptica de las comidas.

Las diferencias sensoriales en las características de calidad de las especias (color, aroma, sabor) se basan principalmente en su origen biológico y geográfico, además muchas son productos de alto costo debido a que provienen de fuentes limitadas, con condiciones de producción y cosecha complicadas, entre éstas se encuentran, el azafrán, la vainilla, la canela, etc. (Reinholds et al., 2015). Aunque hay un amplio rango de países productores de condimentos, China e India son los principales proveedores (CBI, 2015).

Al igual que otros productos agrícolas, las especias pueden estar sujetas a contaminaciones químicas y adulteraciones dentro de uno o más estadios de la cadena de suministro: producción, almacenamiento, transporte y distribución. Además de las pérdidas económicas que incluyen el deterioro de la marca comercial, calidad y ganancias, existen consecuencias internas y externas de posibles peligros para la salud asociados a contaminantes que pueden causar problemas crónicos en la población.

La tabla 1.1 muestra las posibles adulteraciones en especias descritas por American Spice Trade Association (ASTA).

Tabla 1.1. Posibles adulterantes de especias descritas por ASTA

Producto	Adulterante
Orégano	Hojas extrañas ej. sumac, cistus, Hierbas: tomillo, mejorana
Azafrán	Desechos florales, colorantes artificiales
Canela	Cascaras de café
Nuez moscada	Cascaras de café
Especias molidas	Especias agotadas, almidones, granos, cáscaras, oleoresinas
Pimentón	Piel de tomate, dextrosa, monosacáridos, disacáridos, <b>colorantes Sudan, Para-Red y colorantes relacionados</b>
Pimienta blanca y negra molida	Trigo sarraceno y mijo

### 1.2.1 Pimentón

Existen alrededor de 400 especias y hierbas aromáticas comercialmente disponibles, dentro de las especias más consumidas mundialmente se encuentra el pimentón, polvo que se obtiene al moler pimientos rojos secos y que sirve para agregar y/o realzar una gran variedad de preparaciones culinarias, es muy utilizado en Europa, donde el consumo se ha incrementado 1,7% por año entre 2010 y 2013, con un total de 385.000 toneladas en 2012.

El pimiento para pimentón (*Capsicum annuum*), también llamado paprika, es originario del continente americano específicamente de la zona de México, Bolivia y Perú (Figura 1.2).



Figura 1.2. *Capsicum annuum*.

En la actualidad se cultiva en los cinco continentes, siendo los principales productores países asiáticos como India y China, que concentran alrededor del 50% de la producción mundial. El 66% de las importaciones mundiales están lideradas por Estados Unidos, Alemania, Malasia, España, Japón, México y Sri Lanka. A nivel de comercio internacional de especias, en primer lugar, se encuentra la pimienta, luego la mostaza y en tercer lugar se posicionan el pimentón y los ajíes, con una tasa de crecimiento anual del 4%. En nuestro país, el pimiento para pimentón se cultiva hace más de 70 años en las provincias de Salta, Catamarca y Tucumán, convirtiéndose en un cultivo de gran importancia para las economías regionales. El pimiento para pimentón pertenece a la familia de las Solanáceas, como el tomate, el tabaco y la papa. Existen variedades dulces y picantes de *Capsicum annuum L.*, sin embargo, son las dulces las que más se utilizan para la obtención del pimentón. Entre las variedades de pimiento cultivadas en Argentina se encuentran las redondas y las alargadas, tales como Redondo de Cachi, Largo de Cachi, Trompa de Elefante, Tacuil o Santa María, entre otras. Además, en los últimos años se introdujeron variedades provenientes de semilleros reconocidos, entre las que se destacan Papri King, Papri Queen, Jaranda y Mayorca. Recientemente el INTA (Instituto Nacional de Tecnología Agropecuaria), en el marco del Proyecto Nacional de Aromáticas, obtuvo y desarrolló en el Campo Demostrativo de Encalilla en Amaicha del Valle (Tucumán) tres variedades comerciales de *Capsicum*, perfectamente diferenciadas e inscriptas en el Instituto Nacional de Semillas (INASE): “Lautaro” (INTA–2008), “Yokavil” (INTA-2004) y “Encalilla Alargado” (INTA–2012). Estas tienen propiedades interesantes para su comercialización como la “Encalilla Alargado” con un elevado contenido de xantófilas (color >350 ASTA en cáscara), con excelente coloración visual y elevada relación de xantófilas rojas a amarillas, manteniendo un buen nivel de carotenoides pro-vitamina A (Wierna et al., 2015).

Los frutos se cosechan en forma manual entre febrero y abril, en dos o tres recolecciones a medida que alcanzan una coloración rojo intenso, indicador de la madurez de cosecha. En general, el productor realiza también la etapa de secado, siendo ampliamente difundido el secado del pimiento directo al sol. Las vainas de pimiento para pimentón se secan hasta que alcanzan entre 9 y 12% de humedad presentando una consistencia correosa al tacto, esto permite que los frutos se almacenen hasta su molienda sin deteriorarse. Las principales variables que determinan la calidad de secado son la cantidad de calor que recibe el producto, la cantidad de aire seco que permite retirar la humedad y la higiene con que se realiza el proceso. El secado

del pimiento se puede realizar por distintos métodos, la exposición directa al sol sobre el suelo es la más difundida dado su bajo costo (90% de la producción), sin embargo, se obtiene un producto de baja calidad al mantenerse en contacto con el suelo y expuesto a la contaminación (Figura 1.3).



Figura 1.3. Secadero directo al sol.

La alternativa recomendada por el INTA es utilizar tendaleros, colocando los pimientos recién cosechados, sobre una malla de tejido plástico suspendida a 20 cm en estructuras de maderas y alambre, cubiertas con plástico transparente. El material cosechado permanece 15 días en los tendaleros hasta que se seca completamente, con un rendimiento en materia seca entre 18-24%, lográndose entre 90-95% de pimientos de primera calidad limpios de tierra e impurezas (Figura 1.4).



Figura 1.4. Tendaleros para el secado de pimientos.

En la actualidad la provincia de Salta comenzó un proyecto para lograr un producto Premium a través de secadores con paneles solares, cabe resaltar que el funcionamiento de este proceso es sumamente superior, ya que seca en dos días una hectárea de pimentón contra 18 días del secado tradicional.

Para obtener el pimentón, los frutos secos del *Capsicum* se muelen en molinos de piedra o industriales hasta el grado de impalpable. Los molinos de piedra alcanzan temperaturas suficientemente elevadas para causar la evaporación de parte del aceite contenido en el pimiento, impactando negativamente sobre la intensidad tanto del sabor como del color. Desde el punto de vista alimenticio, comercial y tecnológico la coloración roja del pimentón y de su oleorresina es uno de los criterios de calidad fundamental y se expresa por el contenido total de carotenoides presentes en el pimiento. Según el artículo 1233 del Capítulo XVI "Correctivos y Coadyudantes" del Código Alimentario Argentino, con la denominación genérica de Pimentón o Páprika, se entiende el producto obtenido de la molienda de los frutos de Pimiento para pimentón (*Capsicum annuum L.*). El CAA clasifica al pimentón en tres calidades: Extra, Selección y Común, basando esta clasificación en cinco parámetros físico-químicos: humedad, ceniza total, ceniza insoluble, extracto etéreo, fibra bruta y debe encontrarse libre de agregados de aditivos o sustancias extrañas. A partir del año 2008 incorporó el color ASTA, AOAC método 971.26 (Horwitz, 2002), como un parámetro más para su clasificación. La calidad y el precio del pimentón dependen de la intensidad del color, que puede medirse siguiendo el procedimiento establecido por la "American Spice Trade Association" (ASTA). Su determinación consiste en disolver en acetona una muestra de material seco y molido, y medir la absorción de luz monocromática de la solución; estableciéndose una relación entre la absorción de luz y el peso de la muestra se calcula el índice conocido como "grados ASTA". Hoy, este parámetro se ha convertido en el más restrictivo, hasta el punto en que pimentones con valores de color inferiores a 70° ASTA no deberían ser comercializados. A nivel internacional, un pimentón de buena calidad, apto para la exportación, debe superar los 120° ASTA.

Las adulteraciones del pimentón pueden realizarse de diversas maneras, entre otras, por el agregado de harinas de diferentes orígenes como trigo, maíz, arroz o soja, que aumentan el peso, o bien por el agregado de colorantes a fin de mejorar la coloración de los mismos (Araújo Vilar et al., 2014). Desde el año 2003 se ha efectuado en muestras de pimentón comercializadas en establecimientos de Salta, el seguimiento del color, medido en unidades ASTA, así como también de la presencia de adulterantes. Los análisis revelaron un 40% de muestras adulteradas con colorantes naturales

provenientes de semillas de rocú (*Bixa orellana* L.) e identificados como bixina y norbixina (Wierna et al., 2015). Asimismo, se corroboró que la mayoría de las muestras sin adulterar presentaban valores de color inferiores a los establecidos por el CAA. Por consiguiente, siendo el color el atributo fundamental para la evaluación de la calidad de estas especias, el agregado de colorantes constituye la adulteración más frecuentemente.

### 1.3 Colorantes Sudan y Para-Red

Los colorantes Sudan I, II, III y IV y Para Red son una familia de colorantes azoicos sintéticos de intensa tonalidad rojiza y muy estables, que son utilizados ampliamente en la industria como aditivos en ceras, plásticos, tintas de impresión, pinturas y productos para pulir pisos, entre otros. Químicamente los colorantes Sudán I, II y Para-Red son similares entre sí, con la diferencia que el Para-Red contiene un grupo nitro y el Sudán II contiene dos grupos metilos. Esta misma situación se repite en las estructuras de Sudán III y IV que son análogos entre sí (Figura 1.5). Estos colorantes son insolubles en agua y solubles en disolventes orgánicos poco polares.

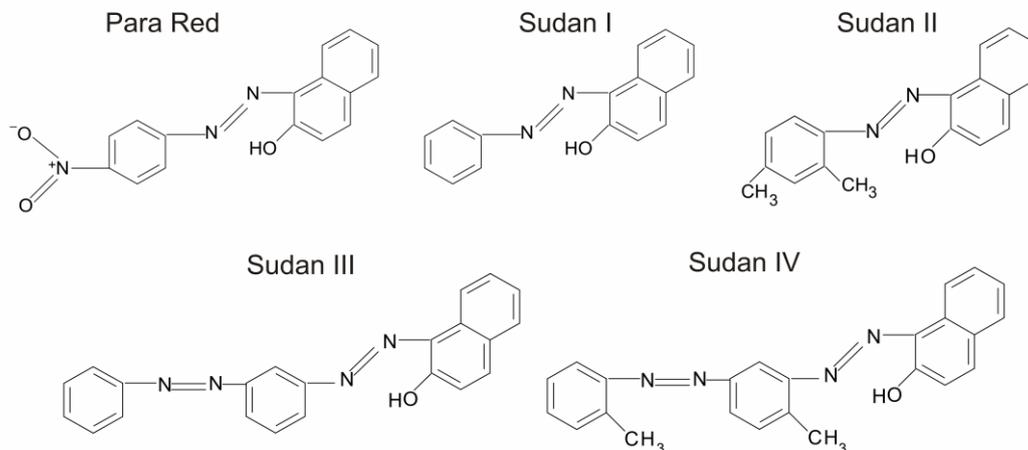


Figura 1.5. Estructura química de colorantes Sudan I-IV y Para Red.

Estos colorantes se han utilizado fraudulentamente en especias culinarias tales como: chile en polvo, pimentón dulce y picante, pimienta de cayena, curry, cúrcuma, etc. con el fin de intensificar y preservar el color de las mismas, aumentando de esta manera su calidad organoléptica. Los mismos se han detectado también, en alimentos elaborados como salsas y aderezos relacionados con estas especias. De acuerdo con la IARC (Agencia de Investigación Internacional del Cáncer) (IARC, 1975), los

colorantes Sudan están considerados carcinógenos de clase 3. Estudios en ratas demostraron que son degradados por enzimas del tipo oxidativo cuando son ingeridos por organismos vivos, generando metabolitos de elevada toxicidad contra el ADN, con probabilidad de inducir mutación genética celular (Xu et al., 2007; Fonovich, 2013). Estos metabolitos biológicamente activos son degradados a aminas aromáticas que pueden actuar directamente sobre células hepáticas, causando un severo daño toxico al hígado, y también puede existir una acción tóxica secundaria en la vejiga por ser la vía de eliminación de los mismos (An et al., 2007).

Debido al grave impacto que provoca en la salud pública la utilización de colorantes Sudan en alimentos, los estados miembros de la Unión Europea se han organizado para implementar medidas reguladoras contra el uso de los mismos. En el año 2003 se reportó la primera alarma de entrada de estos colorantes, principalmente en alimentos provenientes de India, en mayo de ese año el Sistema de Alerta Rápida para Alimentos y Piensos (RASFF) de la Unión Europea notificó por primera vez la entrada de chile picante y productos derivados importados de la India adulterados con Sudán I. Teniendo en cuenta la amenaza para la salud que representaba este colorante fueron decomisados todos los productos sospechosos de estar adulterados. Como consecuencia, los estados miembros de la Unión Europea (UE) establecieron que todas las partidas de chile picante y sus productos derivados (Frutos del género *Capsicum*, secos y triturados o molidos incluidos en el código NC 0904 20 90) importadas debían presentar un informe analítico en el que se demostrara que el producto no contenía colorantes Sudán (Commission Decision, 2003). También se notificó que, en caso de ausencia de un informe analítico, el importador debía someter el producto a los análisis correspondientes, a fin de demostrar que no contenía Sudán I y se pidió a los Estados Miembros que lleven a cabo estrictos controles en los diferentes puntos de venta con el fin de evaluar la magnitud del problema. Este seguimiento reveló la contaminación con Sudán I en una amplia gama de alimentos y productos que incluían chile como ingrediente. En enero del 2004 se adoptó una segunda decisión (Commission Decision, 2004) que extendía la medida hacia otras especias, como curry en polvo y todas las formas del género *Capsicum* como pimentón dulce, picante, ahumado y pimienta de cayena. Esta medida también fue aplicada para los otros colorantes Sudán (Sudán II, III y IV). Finalmente, la cúrcuma y el aceite de palma fueron incluidos en la norma del año 2005, por lo que se ampliaron las medidas de emergencia a estos dos alimentos en su decisión del 23 de mayo de 2005 (Commission Decision, 2005). Por consiguiente, la utilización de colorantes Sudan en alimentos está prohibida en la Unión Europea. La

regulación se reformuló en 2009, estableciéndose la discontinuidad en el requerimiento del reporte analítico de ausencia de colorantes Sudan en alimentos importados, reemplazándolo por estrictos controles en los puntos de entrada a la Comunidad europea (Comisión Decisión, 2009). Por consiguiente, el número de casos de alimentos adulterados con colorantes Sudan disminuyó considerablemente (RASFF, 2012). Sin embargo, actualmente se siguen detectando casos de alimentos adulterados con estos colorantes a elevadas concentraciones, generalmente provenientes de países orientales y principalmente en especias que son vendidas a granel o sueltas en mercados locales y zonas rurales (Tripathi et al., 2007, Mishra et al., 2007). La tabla 1.2 muestra la variedad de alimentos que han sido objeto de adulteración con colorantes de la familia Sudan.

**Tabla 1.2. Alimentos adulterados con colorantes Sudan**

<b>Colorante</b>	<b>Alimento</b>	<b>Referencia</b>
<b>Sudan I</b>	Chile en polvo	Haughey S. et al., (2015)
<b>Sudan I</b>	Chile en polvo, salsa de chile y de tomate	Palanisamy S. et al., (2017)
<b>Sudan I</b>	Chile picante, ketchup	Li B. et al., (2015)
<b>Sudan I</b>	Salsa de chile, de tomate, de fresa y chile en polvo	Mojdeh E. et al., (2013)
<b>Sudan I</b>	Bebidas coloreadas	Gomez M. et al., (2016)
<b>Sudan III</b>	Extracto de paprika	Jahn M. et al., (2015)
<b>Sudan I-IV</b>	Cúrcuma y curry, mariscos, caviar, pescado	Berlina A. et al., (2017)
<b>Sudan I-IV</b>	Curry, pasta de curry, cúrcuma, paprika, chile en polvo y en escamas, salsa de chile	Schummer C. et al., (2013)
<b>Sudan I- mezclas de Sudan I+IV</b>	Paprika, ají molido	Di Anibal C. et al., (2014)
<b>Sudan I-IV</b>	Salsa barbacoa y ketchup	Benmassaouda Y. et al., (2017)
<b>Sudan I-IV</b>	Bebidas	Chailapakul O. et al., (2008)
<b>Sudan I-IV</b>	Salsa de chile picante y de tomate, huevos, salchichas	Zhou M. et al., (2013)
<b>Sudan I-IV</b>	Vino tinto	Song Y. et al., (2015)
<b>Sudan I-IV</b>	Productos origen animal: carne, hígado, piel y huevos (cerdo, pollo, pato y peces)	Chen D. et al., (2013)
<b>Sudan I-IV</b>	Caramelos	Yu W. et al., (2015)
<b>Sudan I-IV y Para- Red</b>	Chile en polvo, salsa de chile, especia nande, salchichas, base de sopa (hotpot)	Li J. et al., (2013)

En Argentina no está permitido el uso de colorantes Sudán en alimentos (CAA, Cap. XVIII) no obstante, el Sudán III está permitido en cosméticos, pero sólo en aquellos que no estén en contacto con las mucosas (Mercosur, 2008). Por consiguiente, sería vital el desarrollo de metodologías analíticas rápidas, simples, económicas y fiables que permitan detectar la presencia de colorantes Sudan en alimentos, con el objetivo de evitar las adulteraciones con las consecuencias que ello tendría en la salud de la

población. A su vez, el desarrollo de estas metodologías sería de gran utilidad en situaciones planteadas dentro del comercio internacional, cuando las partidas de alimentos importados no estén acompañadas de un certificado que demuestre que no contienen colorantes Sudan. En este sentido, es importante la realización de un muestreo al azar, llevado a cabo por las autoridades locales, en puertos y vías de entrada de alimentos importados, para que, en caso de confirmar lotes adulterados los mismos sean inmediatamente decomisados y destruidos.

### 1.3.1 Detección colorantes Sudan y Para-Red

Los métodos comúnmente utilizados para detectar los colorantes Sudan y Para-Red en alimentos están basados en el uso de técnicas cromatográficas (Rebane et al., 2010) con diferentes detectores y tratamientos de muestra (Tabla 1.3). Si bien estas técnicas permiten la detección de Sudán de forma inequívoca y llegan a límites muy bajos, pero presentan la desventaja de requerir elevados tiempos de análisis, instrumentación costosa, procesos de limpieza, pre-concentración y analistas entrenados, etc.

<b>Tabla1.3. Detección de colorantes Sudan y Para-Red en alimentos: Cromatografía</b>				
<b>Analito</b>	<b>Alimento</b>	<b>Técnica</b>	<b>Tratamiento de muestra</b>	<b>Referencia</b>
Para-Red Sudan I-IV	Chile en polvo	HPLC UV-visible	Polímeros de fase interna emulsionada (HIPE)	Du F. et al., (2017)
Sudan I-IV	Ketchup, salsa barbacoa	HPLC UV-visible	Nanocelulosa modificada con nanopartículas de hierro/fase sólida	Benmassaouda Y. et al., (2017)
Sudan I-IV Sudan B, Sudan 7B	Chile en polvo	HPLC UV-visible	Polímeros de impresión molecular de sílica gel modificada	Ji W. et al., (2017)
Sudan I-IV	Jugo	HPLC UV-visible	Líquidos iónicos sólidos/microextracción líquido-líquido microondas	Hu et al., (2016)
Sudan I-IV	Caramelos, bebidas	HPLC UV-visible	Sistema bifásico acuoso de líquido iónico/surfactante aniónico	Yu M. et al., (2015)
Sudan I-III, Sudan B	Aceite de chile	HPLC fluorescencia		Liu J.,Gong Z. et al., (2012)
Sudan I-IV	Carnes y derivados	LC MS/MS	Extracción asistida por microondas/polímeros impresión molecular (MIP-SPE)	Chen D. et al., (2013)
Sudan I-IV	Chile en polvo, snacks de chile	UPLC MS/MS	Extracción líquido/líquido	Schummer C. et al., (2013)
Sudan I-IV	Bebidas	FIA-HPLC-detección electroquímica	Extracción líquido/líquido	Chailapakul O. et al., (2008)
Para-Red, Sudan I-IV, Orange G	Chile en polvo, salsa chile, mezcla de especias	Cromatografía Inyección Secuencial (SIC)	Centrifugación y filtración	Chocholouš P. et al.,(2017)
Sudan I-IV	Chile en polvo	Electroforesis Capilar	Extracción líquido/líquido	Mejia E. et al., (2007)

Alternativamente, se han utilizado otro tipo de técnicas como electroforesis capilar, cromatografía de inyección secuencial, métodos basados en inmunoanálisis o técnicas electroquímicas con una amplia variedad de sensores (Tabla 1.4).

<b>Analito</b>	<b>Alimento</b>	<b>Técnica</b>	<b>Tipo de detección</b>	<b>Referencia</b>
Sudan I	pasta tomate, salsa chile, especias	Inmunoensayo	Inmunoensayo enzimático en modo cinético	Zvereva E. et al., (2016)
Sudan I	salsa chile, ketchup	Electroquímica	Electrodos de película modificada con material conductor	Bang et al., (2015)
Sudan II	salsa chile, ketchup	Electroquímica	Biosensor basado en ADN de esperma de salmón	Ensafi A. et al., (2012)
Sudan I	bebidas coloreadas, colorante líquido comercial	Electroquímica	Voltamperometría de separación por adsorción con electrodo de gota de mercurio	Gómez M. et al., (2016)
Sudan I	chile en polvo	Espectroscópica	Sensor nanopartículas Silicio/quenching fluorescencia	Ammu Rosin J. et al., (2016)

Por otro lado, también se han desarrollado métodos analíticos basados en técnicas espectrométricas como UV-visible, fluorescencia molecular,  $^1\text{H-NMR}$ , Masas, Raman, IR y NIR. Las técnicas espectrométricas tienen la ventaja de generar una rápida respuesta, pero como los alimentos son matrices complejas, comúnmente es necesario la aplicación de tratamientos multivariantes apropiados para obtener información útil de las señales espectrales (Tabla 1.5). De esta manera, la espectrometría con análisis multivariante es de gran utilidad ya que involucra en un análisis cualitativo rápido con instrumentación asequible (espectrómetros y software científico). Asimismo, presentan la ventaja extra de poder realizar análisis *in-situ* mediante su adaptación a equipos portátiles.

**Tabla 1.5 Detección de colorantes Sudan y Para-Red en alimentos: Técnicas espectrométricas con análisis multivariante**

<b>Colorante</b>	<b>Alimento</b>	<b>Técnica</b>	<b>Tratamiento quimiométrico</b>	<b>Referencia</b>
Sudan I, II, III, IV	salsa tomate, aceite palma, chile en polvo	Espectrometría de Masas de alta resolución	Cuantificación univariante	Sciuto S. et al., (2017)
Sudan I, II, III, IV	azafrán	Resonancia Magnética Nuclear ( <sup>1</sup> H-RMN)	Cuantificación univariante	Petrakis E. et al., (2017)
Sudan I, II, III, IV	pimentones	Resonancia Magnética Nuclear ( <sup>1</sup> H-RMN)	Selección de variables, Clasificación multivariante: PLS-DA	Di Anibal C. et al., (2011)
Sudan I, II, III, IV	pimentón, ají molido	UV-visible	Clasificación multivariante sin supervisar y supervisada: PCA, PLS-DA, KNN	Di Anibal C. et al., (2014)
Sudan I	kétchup, salsa barbacoa	UV-visible	Clasificación multivariante (Targeted and Untargeted analysis): PLS-DA, PLS-DM	Di Anibal C. et al., (2016)
Sudan I, II, III, IV	salsa chile y de tomate, huevos, salchichas	Sistema fluorescente basado en Mn-ZnS quantum dots	Cuantificación univariante	Zhou M. et al., (2013)
Sudan I	pimentones	Fluorescencia sincrónica	Clasificación multivariante sin supervisar y supervisada: PCA, PLS-DA	Di Anibal C. et al., (2015)
Sudan I, II, III, IV	snacks de chile	Raman (SERS)	Cuantificación multivariante: PLS	Ou Y. et al., (2017)
Sudan I	pimentón	Raman (SERS)	Clasificación multivariante sin supervisar: PCA	Di Anibal C. et al., (2012)
Sudan I	chile en polvo	NIR / Raman	Cuantificación y clasificación multivariante: PLS / PLS-DA	Haughey S. et al., (2015)
Sudan II	azafrán	FT-IR	Selección de variables, Clasificación multivariante: GA-LDA	Karimi S. et al., (2016)
Sudan I	pimentón	FT-IR	Cuantificación multivariante: Hybrid Lineal Analysis (HLA-GO)	Lohumi S. et al., (2017)
Sudan I	bebidas coloreadas	Imágenes digitales	Cuantificación univariante	Li W. et al., (2016)

#### 1.4 Métodos de screening

En química analítica, los métodos de screening han constituido una herramienta efectiva y de gran importancia a lo largo de los años en muchas áreas y laboratorios de rutina. Se han utilizado ampliamente en ciencias medioambientales, farmacéuticas, medicas, clínicas y en análisis de alimentos, pudiendo ser adaptados a problemáticas específicas (Muñoz-Olivas, 2004). El término “screening” ha sido definido de dos maneras desde el punto de vista analítico: Por un lado, “métodos que indican si los analitos de interés están presentes por encima o debajo de un determinado valor umbral”, por otro lado, “adquisición rápida o semi-cuantitativa de todos los componentes de una muestra”.

En general, estos métodos reúnen las siguientes características (Muñoz-Olivas, 2004):

- Tienden a tener un mayor énfasis cualitativo que cuantitativo.
- La muestra se analiza sin o con mínimo procesamiento.
- Son muy útiles en situaciones que requieren una rápida respuesta.
- La respuesta obtenida a veces requiere confirmación por un método alternativo.

Asimismo, el principal objetivo y las ventajas de los métodos screening se resumen de la siguiente manera:

- Evitar la necesidad de procesar un elevado número de muestras.
- Promover la toma rápida de decisiones.
- Obtener medidas globales del analito.
- Minimizar operaciones preliminares de métodos convencionales que usualmente son tediosos, largos y fuentes de errores sistemáticos, etc.
- Minimizar el uso de instrumentos de elevado costo y mantenimiento, ya que solo las muestras positivas requerirían de los mismos (Muñoz- Olivas, 2004).

Un ejemplo de utilización de estos métodos está constituido por técnicas que proporcionan una respuesta inespecífica que debería estar sujeta a un contexto discriminante. En este sentido, las técnicas espectroscópicas junto con clasificación multivariante han jugado un importante rol en química analítica en los últimos años.

La implementación de metodologías de screening proporciona una herramienta muy útil y práctica para determinar *a priori* el estado de una muestra, que en caso de ser positiva (adulterada) debe ser analizada mediante un análisis confirmatorio, que generalmente involucra instrumentación de mayor sensibilidad y más costosa.

Por consiguiente, estos métodos proveen una solución de gran utilidad en situaciones que requieran una rápida respuesta, permitiendo la toma de decisiones de manera oportuna.

### **1.5 Espectroscopía de Infrarrojo Cercano NIR**

La región del Infrarrojo está comprendida entre 700-10<sup>6</sup> nm, y tanto por razones instrumentales como por las características de interacción de la radiación con la materia, esta región se divide en tres zonas: Infrarrojo cercano (NIR), Infrarrojo medio (MIR) e Infrarrojo lejano (FIR).

Tabla 1.6 Espectroscopía IR

Región	Longitud de onda (nm)	Origen Absorbancia
NIR	700-2500	Sobretonos y bandas de combinación de vibraciones moleculares fundamentales
MIR	2500-50000	Vibraciones moleculares fundamentales
FIR	50000-10 <sup>6</sup>	Rotaciones moleculares

La espectroscopia de Infrarrojo cercano NIR es una técnica vibracional basada en la absorción de radiación electromagnética en un rango de longitudes de onda comprendido entre 700 y 2500 nm, aproximadamente. La absorción en el infrarrojo lejano se debe a rotaciones moleculares, en el infrarrojo medio proviene de vibraciones moleculares fundamentales y el cercano se debe a sobretonos y bandas de combinación de las vibraciones fundamentales de las moléculas. Las intensidades de las bandas de absorción decrecen sucesivamente desde MIR a la región del visible. Los sobretonos ocurren a frecuencias 2 o 3 veces mayores que la de vibración fundamental y la intensidad de las bandas de absorción decrece con el incremento de los sobretonos (Figura 1.6).

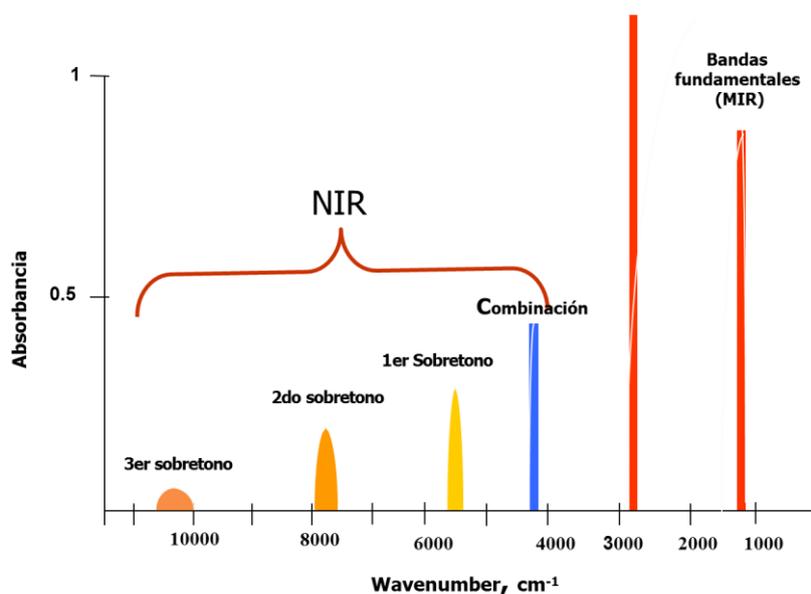


Figura 1.6. Vibraciones fundamentales en NIR.

Las bandas de combinación provienen de la suma de bandas fundamentales de diferentes vibraciones, y se encuentran a menores energías que los sobretonos (Figura 1.7).

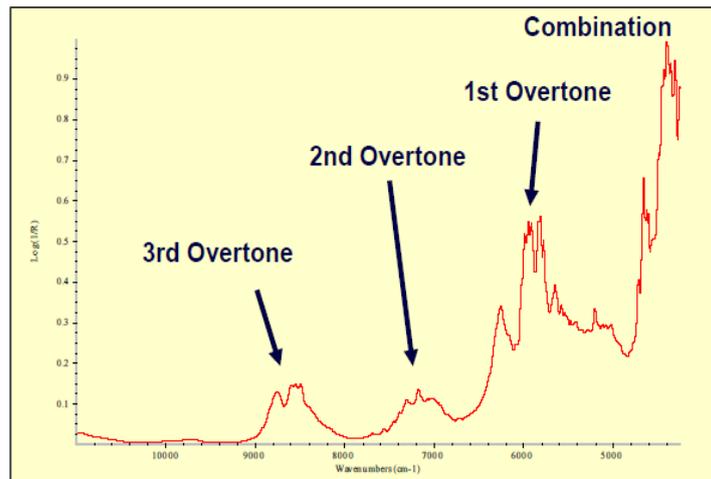


Figura 1.7. Sobretonos y bandas de combinación en un espectro NIR.

Las bandas más frecuentes en NIR son debidas a enlaces que contienen carbono-hidrógeno (CH), oxígeno-hidrógeno (OH) y nitrógeno-hidrógeno (NH). Estos enlaces químicos se encuentran comúnmente en nutrientes como carbohidratos, proteínas, lípidos y también en el agua. Además, la absorción es dependiente de la concentración por lo que NIR puede ser utilizada para cuantificar. En NIR las bandas de absorción no están tan bien definidas como en el infrarrojo medio (MIR), apareciendo en forma de bandas con menor intensidad, anchas y solapadas entre sí, que dan espectros complejos que contienen tanto información útil como irrelevante. Por consiguiente, es necesario un adecuado tratamiento espectral para obtener la información deseada.

Los comienzos de NIR datan del año 1881, pero las primeras aplicaciones analíticas se desarrollaron en la década del 50 en el siglo XX, con la aparición de los primeros espectrofotómetros comerciales. El primer impulso importante fue dado en la década del 60 cuando se utilizó en el estudio de matrices complejas de origen vegetal orientado al campo agroalimentario, esto despertó aún más el interés en esta técnica. Desde entonces, NIR ha sido ampliamente aplicada al campo de los alimentos y la agricultura, fundamentalmente para el control de calidad en granos, cereales y semillas. Es utilizada como rutina para en análisis composicional, funcional y sensorial de ingredientes alimentarios, tanto en procesos intermedios como en productos finales.

Hoy en día, es una las técnicas de huellas dactilares ampliamente utilizada como análisis de screening, con resultados exitosos para la detección de adulteraciones alimentarias como así también en el control de calidad en la industria, su rapidez permite realizar medidas frecuentes desde la recolección de materias primas, producción,

control del proceso hasta la calidad del producto final (Burns y Ciurczak, 2008). En el campo de especias y hierbas, NIR ha sido empleado para análisis cuantitativo y autenticación (Tabla 1.7).

**Tabla 1.7 NIR en especias**

Especia	Analito	Modo medida	Referencia
Aceite tomillo, orégano y manzanilla	Carvacrol, thymol and $\alpha$ -bisabolol	Reflectancia	Schulz H. et al., (2007)
Pimentón	Micotoxinas	Reflectancia	Hernández-Hierro J. et al.,(2008)
Cúrcuma	Curcuminoides	Reflectancia	Tanaka K. et al., (2008)
Polvo ginseng	Ginsenoides	Reflectancia	Ren G. et al., (1999)
Azafrán	Control calidad/glicósidos de crocetina	Reflectancia	Zalacain A. et al., (2005)

Existen tres tipos de medida en NIR: transmitancia, reflectancia y transflectancia. En el modo transmitancia la luz atraviesa la muestra y llega al detector, se pueden medir gases, líquidos, semilíquidos y sólidos, en el modo reflectancia el haz de luz es reflejado por la muestra y recogido por el detector, se pueden medir semisólidos y sólidos y en modo transflectancia el haz de luz atraviesa la muestra, se refleja en un reflector que está en contacto con la misma y finalmente llega al detector, se aplica a muestras líquidas y semilíquidas. La figura 1.8 muestra las tres configuraciones instrumentales de los modos de medida NIR.

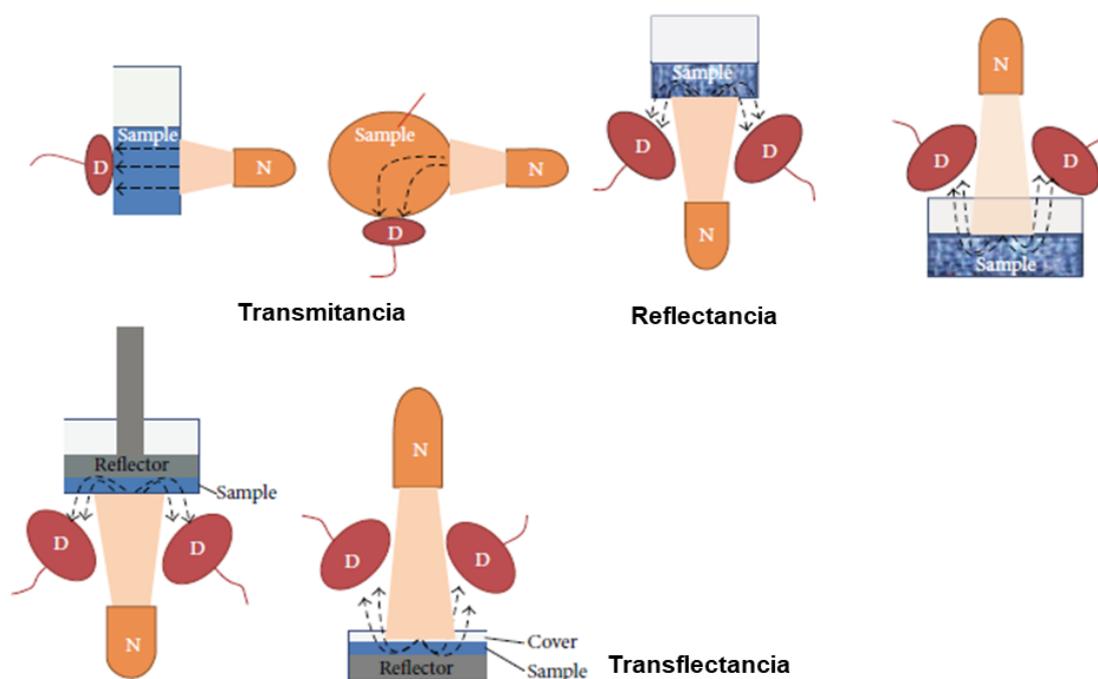


Figura 1.8. Configuraciones instrumentales de medida NIR.

Las técnicas espectroscópicas como NIR generan una gran cantidad de datos relativos a las muestras analizadas en breves intervalos de tiempo. Por otro lado, las bandas de absorción en NIR son generalmente anchas y se superponen generando espectros complejos que contienen tanto información útil como a veces irrelevante. En el caso de muestras sólidas, los espectros NIR pueden estar significativamente influenciados por variaciones sistemáticas no deseadas causadas principalmente por dispersión de la luz; como consecuencia de la variación del tamaño de partícula y diferencias en el camino óptico efectivo, provocando desplazamientos en la línea de base y efectos de no linealidad (Rinnan et al., 2009). Es por ello que al trabajar con espectros NIR se requiere del uso de herramientas matemáticas para el tratamiento de los datos. En este contexto, la quimiometría permite extraer la información presente en los espectros para tener información fidedigna y una correcta interpretación del sistema en estudio.

Entre las ventajas que posee la espectroscopia NIR se puede destacar: versatilidad y adaptabilidad para analizar muestras de diferente naturaleza, técnica no destructiva que permite medir la muestra en su estado original sin tratamiento previo, la no utilización de reactivos químicos. Por otra parte, NIR ha evolucionado instrumentalmente gracias al avance en el campo de la electrónica y óptica, desarrollando equipos FT-NIR donde se pueden realizar medidas muy rápidas y altamente reproducibles, convirtiéndose en una técnica muy útil y versátil con un amplio margen de aplicabilidad. El desarrollo de nuevas herramientas quimiométricas para el tratamiento de datos y la miniaturización de componentes instrumentales que facilitan la portabilidad de los equipos hacen que esta técnica esté en plena fase de auge y expansión (Mercader, 2009).

## **1.6 Análisis multivariante**

La quimiometría es la disciplina química que usa métodos matemático-estadísticos para diseñar o seleccionar los procedimientos óptimos de medida y experimentales, y proveer la máxima información química a partir de análisis de los datos obtenidos (Otto, 2017). Esta disciplina ofrece diversas herramientas para distintos fines como filtrado de señales (pretratamientos aplicados a los datos experimentales para mejorar la calidad de la señal), técnicas de reconocimiento de patrones (establecimiento de agrupaciones de muestras en función de su similitud con el fin de clasificarlas) y de calibración

(establecimiento de una relación cuantitativa entre la señal analítica y alguna propiedad de la muestra) (Vandeguiste et al., 1998).

### 1.6.1 Pre-procesamiento de espectros

El preprocesado de espectros es el paso más importante antes de aplicar un modelo multivariante, ya sea un análisis exploratorio de datos o un modelo de calibración o clasificación. Consiste en correcciones basadas en tratamientos matemáticos con la finalidad de aumentar la calidad de la señal instrumental, reduciendo al máximo la información no asociada a la propiedad estudiada, como pueden ser desplazamientos de la línea base, ruido espectral, etc. Se denomina ruido al componente espectral no relacionado con ninguna propiedad física o química de la muestra como puede ser el ruido instrumental o de iluminación. Además, otras fuentes de variación como la temperatura y humedad pueden aportar interferencias a la señal analítica. En la literatura se han propuesto varios algoritmos matemáticos utilizados comúnmente en NIR, estos pueden dividirse en dos grupos: métodos de corrección de scattering (MSC, SNV, detrending y Wavelets) y derivadas espectrales (Rinnan et al., 2009).

#### Métodos de corrección de scattering

- *MSC: Multiplicative Scatter Correction*

La transformación MSC (Figura 1.9) se basa en el hecho que, para un conjunto de muestras similares, el valor de la absorbancia está linealmente relacionado con la media de los valores espectrales. El primer paso consiste en establecer un modelo lineal con los coeficientes de corrección teniendo en cuenta el espectro promedio:

$$\mathbf{x}_{org} = \mathbf{b}_0 + \mathbf{b}_{ref} * \mathbf{x}_{ref} + \mathbf{e}$$

donde  $X_{org}$  es el espectro original,  $X_{ref}$  es el espectro promedio y  $b$  son los coeficientes de regresión. El espectro corregido ( $x_{corr}$ ) se obtiene como:

$$\mathbf{x}_{corr} = \mathbf{x}_{org} - \mathbf{b}_0 / \mathbf{b}_{ref}$$

- *SNV: Standard Normal Variate*

La transformación SNV es similar al MSC y se define como:

$$\mathbf{x}_{\text{corr}} = \mathbf{x}_{\text{org}} - \mathbf{a}_0 / \mathbf{a}_1$$

donde  $a_0$  y  $a_1$  son el promedio y la desviación estándar del espectro a ser corregido. La principal diferencia entre ambos métodos es que el SNV se aplica a cada espectro independientemente, mientras que en MSC es necesario uno de referencia (Figura 1.9).

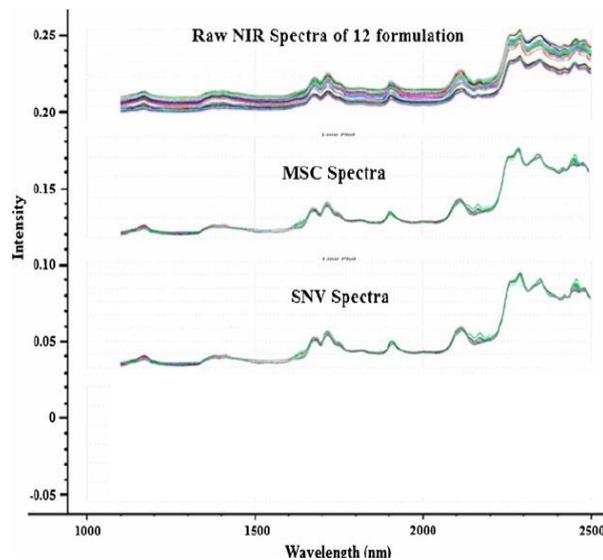


Figura 1.9. Espectros pre-procesados con SNV y MSC.

- *Detrending (Offset)*

Este método consiste en aproximar la línea de base del espectro con una función polinómica, que se resta posteriormente para obtener el espectro corregido. Cuando se emplea un polinomio de grado 0 se denomina *offset correction* y consiste en restar a cada espectro la absorbancia medida a una determinada longitud de onda (zona sin absorción) (Figura 1.10).

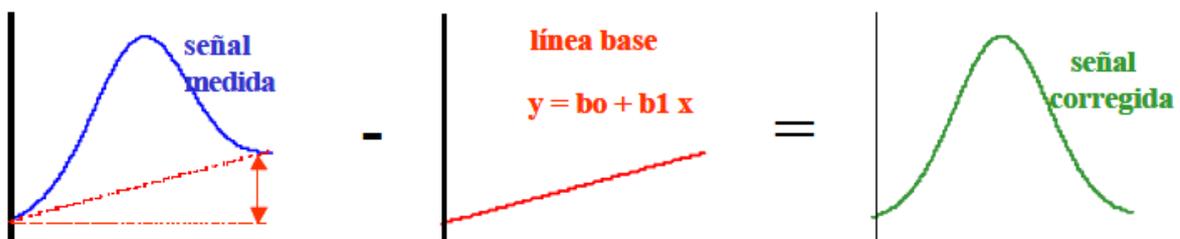


Figura 1.10. Esquema de corrección de la línea base utilizando un modelo explícito lineal.

- *Wavelets*

La transformada wavelet se ha utilizado para comprimir datos, para seleccionar variables y para eliminar ruido y/o background. Wavelets descompone la señal espectral en bandas de baja y alta frecuencia, proporcionando información en el dominio tiempo y frecuencia, con lo cual es posible distinguir entre bandas espectrales anchas y angostas seleccionando coeficientes wavelets a diferentes niveles, los cuales se relacionan con las absorciones de interés (Ge et al., 2007).

### Derivadas espectrales

El proceso de derivación realza la información espectral permitiendo una mayor diferenciación de picos solapados, como también elimina desplazamientos lineales y cuadráticos de la línea base (Figura 1.11). La derivada reduce la correlación entre variables y el efecto de la dispersión debido al tamaño de las partículas. Presenta el inconveniente de magnificar el ruido en la señal, por lo que se utiliza en combinación con métodos de suavizado de la señal como el del ajuste polinómico móvil (método Savitsky-Golay). No suelen utilizarse derivadas superiores a orden 2 (segunda derivada), ya que decrece la relación señal-ruido (S/N).

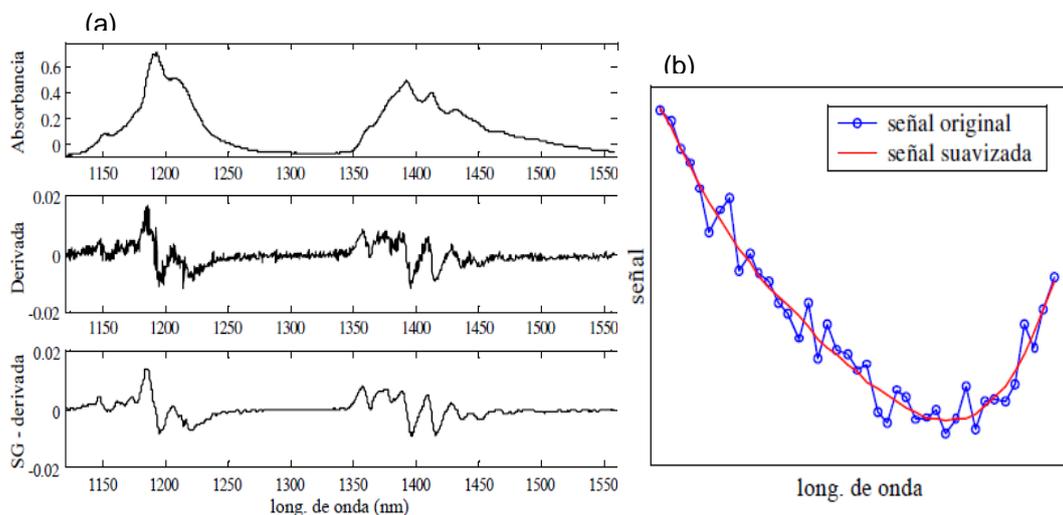


Figura 1.11. (a) Segunda derivada de un espectro sin suavizado (medio) y con suavizado Savitsky-Golay (inferior), (b) Señal suavizada con una ventana de 21 puntos y un polinomio de tercer grado.

## 1.6.2 Técnicas de reconocimiento de patrones

Las técnicas de reconocimiento de patrones pertenecen a una de las áreas de la quimiometría muy utilizadas desde hace décadas. Su filosofía está basada en la identificación de relaciones y/o vínculos entre objetos previamente caracterizados a través de medidas, con el fin de agruparlos y clasificarlos según la similitud que exista entre ellos. Desde un punto de vista geométrico, esto corresponde a identificar regiones en el hiperespacio de las variables que corresponden a diferentes clases. En este contexto, el término clase o categoría indica la colección de objetos que comparten características similares cuya definición dependen del problema en estudio, de manera que el mismo set de muestras puede ser agrupado de distintas maneras de acuerdo con finalidad del mismo. Las técnicas de reconocimiento de patrones se pueden dividir en dos grandes grupos: **no supervisadas y supervisadas** (Brereton, 2009).

### 1.6.2.1 Técnicas de reconocimiento de patrones no supervisadas

Las técnicas **no supervisadas** no poseen información previa de la clase a la que pertenecen los objetos, con el fin de determinar la estructura existente en el conjunto de los datos y encontrar entre ellos grupos de composición o comportamiento similar. Entre ellas, las más utilizadas son el **Análisis por Componentes Principales (PCA)** y las **técnicas de clustering** (agrupamiento).

El **Análisis de componentes principales (PCA)** es una técnica de descomposición/proyección bilineal capaz de condensar una gran cantidad de información en unas pocas componentes principales, comprime los datos mediante unas pocas componentes principales (PCs) que son combinaciones lineales de las variables originales. Las PCs son ortogonales entre sí (Figura 1.12), por lo que cada una provee información no contenida en la otra y en forma decreciente (PC1 recoge la máxima información).

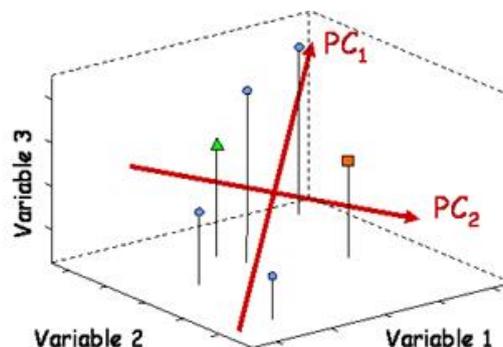


Figura 1.12. Representación de las componentes principales (PCs) en un espacio tridimensional.

Gráficamente PCA proporciona las relaciones que existen entre las distintas muestras a través de un gráfico de scores donde se puede observar la discriminación entre los distintos grupos. También se pueden visualizar las relaciones entre las variables a través de un gráfico de loadings y establecer su relevancia en el modelo. Tiene la ventaja de proporcionar un gráfico biplot que permite ver simultáneamente las muestras y las variables, esto es muy útil para establecer relaciones entre muestras, entre variables y entre muestras y variables.

Las **técnicas de clustering** se basan en encontrar agrupamiento dentro de los datos mediante diferentes caminos para definir proximidad (ya sea similitud o diferencia), encontrando objetos similares como representaciones de puntos cercanos en el espacio de variables que los describen. La similitud/disimilitud entre muestras se obtiene a través de medidas de distancias (por ejemplo, la distancia Euclídea). Las **técnicas de clustering** se dividen en dos grupos: **particionadas** (segmentan el conjunto de datos en k clusters) y **jerárquicas** (descomponen los datos en varios niveles de similitud/disimilitud). Estas técnicas pueden ser utilizadas como complementarias a las técnicas de proyección PCA.

#### 1.6.2.2 Técnicas de reconocimiento de patrones supervisadas

Las técnicas de **reconocimiento de patrones supervisadas** utilizan la información de las clases para establecer un modelo de clasificación a priori. En efecto, una vez que las clases son definidas, estas técnicas de clasificación utilizan los datos de entrenamiento (training set: objetos cuya verdadera clase es conocida) para construir una regla de clasificación que permita clasificar los objetos desconocidos en las clases previamente designadas. Existen en literatura numerosos algoritmos de clasificación de diferente naturaleza, con distintas propiedades y habilidades, ofreciendo una amplia variedad de opciones para resolver problemas de clasificación (Marini, 2013). Por un lado, se pueden dividir en dos grandes grupos de acuerdo a la forma matemática de establecer los límites entre las distintas clases: **técnicas lineales y no lineales**. Las primeras construyen modelos a través del mejor límite lineal para discriminar entre las clases mientras que las segundas se basan en establecer la mejor curva (límite no lineal cuadrático o polinómico) para separar las clases. Las técnicas de clasificación pueden ser **paramétricas**, en las cuales se asume una distribución normal de los datos, o **no paramétricas**, en las cuales no se hace una estimación explícita de la distribución que siguen los datos. Finalmente, otra importante distinción entre las técnicas de

clasificación en base a si están focalizadas en discriminar entre las diferentes clases (**discriminantes**) o si están dirigidas a **modelar las clases**. Las **técnicas discriminantes** se construyen maximizando las diferencias entre las clases, dividiendo el espacio multidimensional en tantas regiones como número de clases en estudio, de esta manera una muestra siempre será clasificada a una de las clases disponibles (Figura 1.13 a). Por otro lado, las técnicas de **modelado de clases** apuntan a capturar la similitud entre las muestras de una misma clase en vez de las diferencias entre las distintas clases. Desde el punto de vista algorítmico cada clase es modelada individual e independientemente de las otras, y desde el geométrico equivale a identificar un volumen en el espacio multidimensional que encierra la clase, de manera que si una muestra cae dentro del mismo es aceptada por esa clase mientras que si cae fuera es rechazada por ese modelo de clase (Figura 1.13 b). En este caso existen varias posibilidades de clasificación: muestras que son aceptadas por una clase, muestras que son aceptadas por más de una clase y muestras que no son aceptadas por ninguna de las clases disponibles.

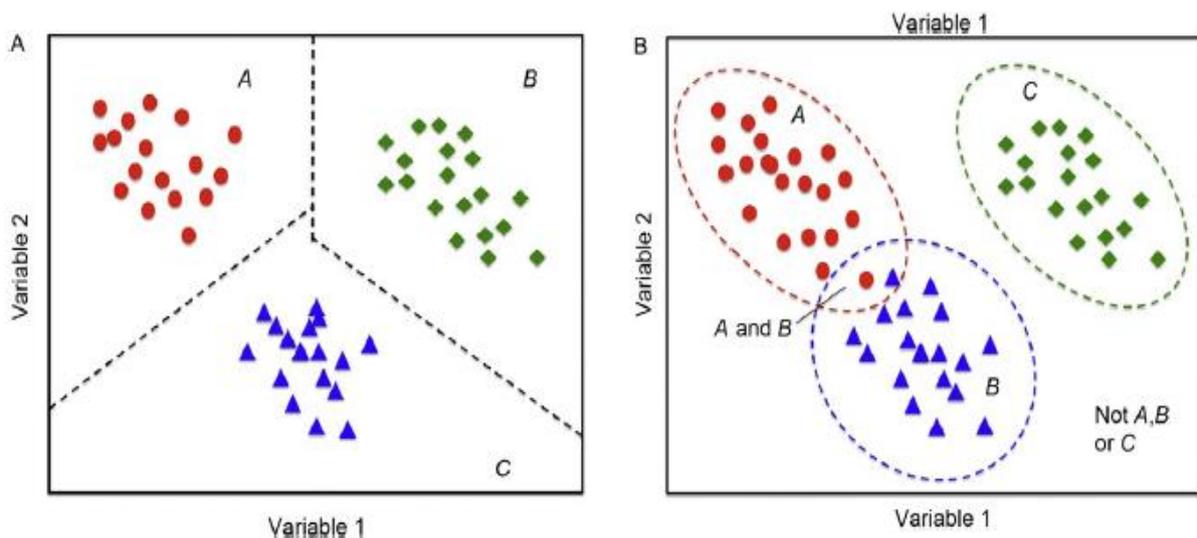


Figura 1.13. Representación de técnicas discriminantes (a) y de modelado de clase (b) en un problema de clasificación de tres clases en un espacio bidimensional.

Ejemplos de técnicas de clasificación **discriminantes** son Linear Discriminant Analysis (LDA), Partial Least Squares Discriminant Analysis (PLS-DA), Quadratic discriminant analysis (QDA), K-Nearest Neighbours (KNN), Support Vector Machines (SVM) y dentro de las técnicas de **modelado de clases** están Soft Independent Modelling of Class Analogy (SIMCA) y Unequal Class-Modelling (UNEQ). La elección

de la técnica de clasificación debe ser hecha en base a la característica de los datos y el objetivo del análisis.

**El Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA)** se basa en el modelo de regresión aplicado a un contexto de clasificación, cuya diferencia radica en la matriz Y (en regresión concentraciones y en clasificación clases), con el fin de reducir la dimensión de los datos y establecer límites entre las clases. En PLS, la varianza de la matriz X (respuestas instrumentales) e Y se descompone por estimaciones sucesivas de componentes PLS que capturan la varianza y correlación entre X e Y. Se crea un modelo donde se maximiza la correlación entre X e Y mediante variables latentes (LVs), donde la identificación de pertenencia a las clases se realiza a través de una matriz de variables ficticias (bloque Y) que consiste en 1 y 0 (Tabla 1.8). El modelo PLS-DA construido se utiliza para predecir nuevas muestras, en donde a cada clase pertenecerán las muestras cuyos valores de Y estén por encima del valor delimitado (umbral) correspondiente a esa clase.

**Tabla 1.8. Identificación de clases en PLS-DA con y sin adulterante**

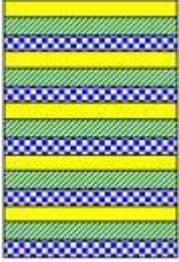
	Clase	Bloque X	Bloque Y					
1	Pimentón O.		1	0	0	0	0	0
2	Pimentón PR		0	1	0	0	0	0
3	Pimentón SI	Datos	0	0	1	0	0	0
4	Pimentón SII	NIR	0	0	0	1	0	0
5	Pimentón SIII		0	0	0	0	1	0
6	Pimentón SIV		0	0	0	0	0	1

O: Original; PR: Para-Red; S I-IV: Sudan I, II, III y IV.

### 1.6.3 Validación cruzada

A veces no se cuenta con la cantidad suficiente de muestras para dividir al conjunto de datos en entrenamiento y prueba, ya sea porque el análisis es costoso, la obtención de datos resulta complicada experimentalmente o porque no es posible armar un conjunto suficiente de muestras. En estos casos, existen estrategias de validación cruzada en las cuales se divide al conjunto de datos en entrenamiento y prueba sistemáticamente **Validación cruzada o cross-validation** (Tabla 1.9).

Tabla 1.9. Esquemas de validación cruzada

	Persianas venecianas	Bloques contiguos	Subconjuntos aleatorios	Uno por vez
Esquema de selección de muestras prueba				
Número experimentos de subvalidación	S	s	(s * r)	N
Grupos de cancelación	N / s	N / s	N / s	1

s: número de divisiones de datos, N: número total de muestras, r: número de iteraciones

El conjunto de calibración inicial se divide en diferentes grupos de cancelación. Cada vez que se calcula el modelo se deja un grupo de cancelación fuera y se calcula un modelo con los demás datos. Una vez realizada la calibración se prueba el modelo con los datos del grupo de cancelación excluido. Se repite el proceso de manera sucesiva dejando cada vez un grupo de cancelación fuera del proceso de cálculo. Si el número de grupos de cancelación es igual al número de objetos en el conjunto de calibración, cada vez quedará un objeto fuera del cálculo del modelo y el proceso se denomina **full-cross-validación o leave-one-out**.

#### 1.6.4 Selección de variables

En literatura, existe una gran variedad de algoritmos basados en diferentes criterios para seleccionar variables (Zou et al., 2010), con el fin de mejorar los resultados de clasificación obtenidos. El objetivo de las técnicas de selección de variables es capturar aquellas del set de datos original que están más específicamente relacionadas con el problema en estudio, y excluir aquellas que están afectadas por otras fuentes de variación y no proporcionan información relevante. En este contexto, se prefiere retener las variables que posean la máxima capacidad discriminante entre las clases. Las ventajas que ofrece el trabajar con un set de variables reducidas son: mejorar la habilidad y robustez de los modelos multivariantes, prevenir el sobreajuste y tener una

mejor interpretación de los datos. Dentro de las variantes, se encuentran métodos basados en intervalos (i-modelos).

### **Interval PCA (iPCA)**

Este método divide el conjunto total de datos en intervalos regulares de igual tamaño y adyacentes, calculando un modelo PCA en cada uno, por lo que se puede obtener información adicional respecto a la principal fuente de varianza de las diferentes regiones espectrales. Los resultados se presentan en múltiples gráficos de scores; siendo útil para visualizar agrupamientos y patrones entre las muestras. De esta manera se puede escoger el/los intervalos que proporcionen la mejor discriminación entre las clases en estudio (Marini, 2013).

### **Successive Projections Algorithm-Linear Discriminant Analysis (SPA-LDA)**

Este algoritmo combina una técnica de selección de variables Sucessive Projections Algorithm (SPA) y clasificación discriminante Linear Discriminant Analysis (LDA). SPA se basa en una secuencia de operaciones de proyección de la matriz de respuestas instrumentales. Inicialmente se construye una cadena ordenada de subconjuntos de variables (variables candidatos), donde cada elemento es seleccionado en base a la menor colinealidad con respecto a las variables vecinas. Luego, los subconjuntos de variables candidatos se evalúan en base a una función de costo relacionada con el promedio de errores de clasificación de un set de muestras utilizado para validación. Una vez que las variables han sido seleccionadas, la clasificación de nuevas muestras se realiza calculando la distancia Mahalanóbica de estas muestras con respecto al vector medio de cada clase (calculado sobre los datos de entrenamiento con las variables seleccionadas), y las muestras son asignadas a la clase cuya distancia es menor (Carreiro Soares et al., 2013).

### **1.6.5 Evaluación de la metodología desarrollada**

La validación de un modelo se considera como el proceso para asegurar que un procedimiento analítico sea fiable y pueda satisfacer las expectativas de una aplicación particular. En resumen, significa que se puede utilizar con confianza.



Figura 1.14. Validación y Evaluación de la habilidad de las metodologías desarrolladas

El desarrollo y la validación de un método están estrechamente relacionados, ya que los parámetros de calidad a menudo se evalúan como parte del desarrollo del método. Al abordar un problema analítico, se debe tener en cuenta consideraciones como se esquematizan en la figura 1.14. La solución del problema analítico debe considerarse como un proceso cíclico e interactivo de verificación y evaluación del método, que no se detiene hasta que el método se considere capaz de satisfacer los requisitos. El proceso comienza con el estudio del problema, lo que se sabe sobre él y cuáles son los requisitos analíticos, de manera de elegir el método que mejor responda, y si se ajusta al propósito, éste debe ser validado mediante la evaluación de los parámetros de calidad del modelo estudiado.

La habilidad de las metodologías desarrolladas se evalúa en términos de los resultados de clasificación obtenidos, en cuyo caso se pueden obtener asignaciones correctas y errores de clasificación. Las asignaciones correctas comprenden los aciertos positivos (**true positives, TP**) que son las muestras clasificadas correctamente a su clase y los aciertos negativos (**true negatives TN**) que son muestras pertenecientes a otra clase e identificadas como externas correctamente. En cambio, dentro de los errores de clasificación tenemos los **falsos negativos (FN)** que son muestras de una clase que no son reconocidas como tal y **falsos positivos (FP)** que son muestras de otra clase incorrectamente identificadas como pertenecientes a la clase.

Los parámetros de calidad son un conjunto de atributos medibles que definen la calidad de un método analítico, que, dependiendo de la naturaleza del problema, del método propuesto y del nivel de validación, pueden ser los siguientes: **sensibilidad**,

**especificidad, precisión, eficiencia del modelo, coeficiente de correlación de Mathews, curvas ROC (Receiver Operating Characteristic Curves), área bajo la curva (AUC)** (Oliveri y Downey, 2012).

**Sensibilidad =  $TP / TP + FN$**  representa la fracción de muestras pertenecientes a una clase que son correctamente aceptadas.

**Especificidad =  $TN / TN + FP$**  representa la fracción de muestras que no pertenecen a una clase y que han sido correctamente rechazadas.

**Precisión =  $TP / TP+FP$**  indica la relación entre el número de muestras correctamente aceptadas y el número total de muestras aceptadas por una clase.

**Eficiencia:  $TP+TN / TP+TN+FP+FN$**  representa la media geométrica de la sensibilidad y especificidad. Puede variar entre 0 (parámetros de sensibilidad o especificidad nula) y 1 (ambos parámetros tienen el máximo valor de 1).

$$\text{Coeficiente correlación Matthews} = \frac{TP \cdot TN - FP \cdot FN}{[(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)]^{1/2}}$$

Este coeficiente tiene en cuenta todos los posibles resultados a obtener. Puede variar entre +1 (todas las muestras perfectamente asignadas) y -1 (todas las muestras más clasificadas) y 0 en el caso de asignación azarosa de muestras.

**Curvas ROC:** Se obtienen graficando la sensibilidad vs. (1 - especificidad) variando alguno de los parámetros de clasificación (límite establecido entre clases, nivel de confianza, etc. que dependerá de la técnica de clasificación utilizada). Como se muestra en la Figura 1.15, una línea diagonal desde el extremo inferior izquierdo hasta el extremo superior derecho representa modelos que proporcionan asignación de muestras al azar (i) mientras que curvas que se ubican por encima de esta línea y tienden hacia el extremo superior izquierdo indican modelos eficientes (iii), pudiendo haber casos intermedios (ii). Estas curvas pueden ser utilizadas para comparar diferentes modelos y para identificar el óptimo compromiso entre sensibilidad y especificidad para un determinado modelo.

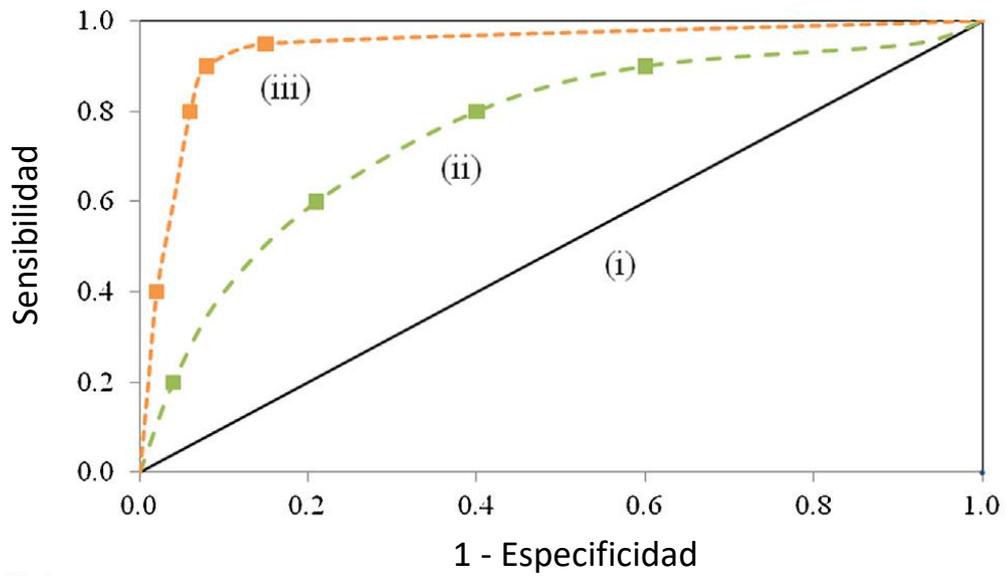


Figura 1.15. Curvas ROC.

## **2. OBJETIVOS**

## 2.1 Objetivos

El objetivo de esta tesis es desarrollar una metodología analítica de *screening* para la determinación de la adulteración de alimentos con colorantes no permitidos, basada en el uso de técnicas espectroscópicas y análisis multivariante.

*Los objetivos particulares son:*

- Determinar la adulteración de pimentones con colorantes no permitidos
- Los colorantes estudiados son Sudan I, II, III, IV y Para Red
- Utilizar técnicas espectroscópicas: Infrarrojo Cercano (NIR).
- Realizar un pre-procesamiento de datos para obtener información útil, relevante y compensar diferencias no espectrales.
- Construir un conjunto de datos que incluya muestras originales (sin adulterar) y adulteradas en las concentraciones de estudio de los mismos.
- Implementar análisis exploratorio de datos con técnicas de reconocimiento de patrones no supervisado (Análisis por Componentes Principales PCA).
- Construir modelos de clasificación con técnicas de reconocimiento de patrón supervisadas (Análisis Discriminante por Mínimos Cuadrados Parciales PLS-DA).
- Mejorar resultados de clasificación obtenidos mediante técnicas de selección de variables (i-PCA y Succesive Projections Algorithm-Linear Discriminant Analysis SPA-LDA).
- Validar las metodologías desarrolladas.
- Postular la metodología propuesta como método de *screening*.

### 3. MATERIALES Y MÉTODOS

### **3.1 Muestras y adulterantes**

Las muestras de pimentones fueron adquiridas en diferentes establecimientos comerciales de la localidad de Bahía Blanca. Las variedades disponibles fueron Pimentón dulce, Pimentón picante, Pimentón dulce ahumado, Pimentón picante ahumado y Pimentón ahumado.

Los colorantes Sudan I, II, III, IV y Para-Red fueron adquiridos de Sigma-Aldrich (St. Louis, MO, USA).

### **3.2 Preparación de muestras**

Se pesaron 2 g de pimentón en balanza analítica en todas sus variedades. Las muestras adulteradas se prepararon a partir del agregado de cantidades adecuadas de cada colorante en los tres niveles de concentración estudiados: 1; 2,5 y 5% (p/p). Las muestras se procesaron durante 10 minutos con un tamiz #70 (American Society for Testing Materials, ASTM) para obtener un tamaño de partícula uniforme, corroborándose la correcta homogenización mediante un estereomicroscopio Olympus trinocular SZ-PT con cámara de video Sony 151A. Posteriormente las muestras fueron conservadas en desecador.

### **3.3 Instrumentación y software**

Los espectros NIR se registraron en un espectrómetro Thermo Scientific Nicolet iS50 FT-IR con módulo NIR (Figura 2.1), equipado con un detector InGaAs, con material de referencia interna de oro (alta reflectancia difusa) y con filtro de CaF<sub>2</sub>, operando en un rango espectral de 12.000 a 4000 cm<sup>-1</sup>.



Figura 2.1. Instrumento NIR

El esquema del instrumento NIR utilizado (Figura 2.2) se compone de una fuente de radiación, que es una lámpara halógena de filamento de tungsteno con ventana de cuarzo. En instrumentos no dispersivos se utiliza la transformada de Fourier (FT) que trabaja dividiendo la luz en dos haces cuyas longitudes de trayectoria se pueden variar periódicamente para dar modelos de interferencia mediante un espejo móvil. Variando la posición de estos espejos se pueden crear interferencias constructivas y destructivas. Recombinando estos haces se obtiene un interferograma (representación de la señal en dominio tiempo) el cual mediante la transformada de Fourier se convierte en el espectro (representación en el dominio de frecuencias). Por último, el detector de Arseniuro de Indio y Galio (600 y 1900 nm) es un material semiconductor para realizar medidas por transmisión en sólidos.

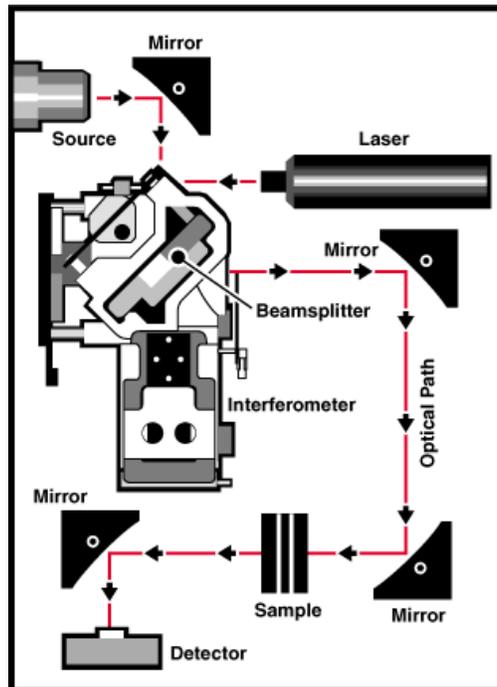


Figura 2.2. Esquema instrumento NIR

Las muestras fueron analizadas en modo de reflectancia difusa por duplicado en viales de vidrio de 1,8 cm de diámetro considerando el mismo nivel de llenado para todas las muestras. Las condiciones instrumentales son:

- Resolución de  $8 \text{ cm}^{-1}$
- 64 scans por muestra
- Modo esfera integradora
- Rango de trabajo de  $10.000$  a  $4.000 \text{ cm}^{-1}$
- Gain 1,0
- Material de referencia de teflón como background externo

Los correspondientes espectros se obtuvieron con el software OMNIC y se exportaron en formato csv y posteriormente se procesaron mediante el software científico Matlab (Versión R2010a).

### **3.4 Análisis multivariante**

El procesado de espectros se realizó con los siguientes algoritmos: Multiplicative Scattering Correction (MSC), Standard Normal Variate (SNV), corrección polinómica de línea de base, primera y segunda derivada y Wavelets. Seguidamente se implementó un análisis exploratorio de datos con Análisis por Componentes Principales (PCA) y se construyeron modelos de clasificación multivariante con Análisis Discriminante por Mínimos Cuadrados Parciales (PLS-DA). Posteriormente se implementaron estrategias de selección de variables con el fin de mejorar los resultados de clasificación mediante modelos basados en intervalos (i-modelos) y el algoritmo de proyecciones sucesivas con análisis discriminante (SPA-LDA).

## 4. RESULTADOS Y DISCUSIÓN

#### 4.1 Tratamiento de muestras

Las muestras de pimentón fueron adquiridas en diferentes comercios locales. Para el estudio se utilizaron las variedades dulce, picante y ahumado, nacionales e importadas, tanto envasadas como comercializadas a granel (Tabla 3.1).

**Tabla 3.1. Pimentones**

Especia	Marca	Comercio adquisición	fecha vencimiento	ORIGEN
PD	LA PARMESANA	WAL-MART	Enero 2017	ARG
PD	ALICANTE	WAL-MART	Junio 2015	ARG
PD	NUESTRAS ESPECIAS	WAL-MART	Nov. 2016	ARG
PD	GATO NEGRO	GATO NEGRO	Enero 2017	ESPAÑA
PD	DUSEN	MERCADO BAHIA	Enero 2017	ARG
PD	HOJAS VERDES	MERCADO BAHIA	Enero 2017	ARG
PD	PRIMER PRECIO	HIPER COOPERATIVA	Dic. 2016	ARG
PD	DON UBALDO	CONTE-GRAND	Dic. 2016	ARG
PD	BEIRAMAR	CONTE-GRAND	Dic. 2016	ARG
PD	KO-KI-TO	CONTE-GRAND	Dic. 2016	ARG
PP	NUESTRAS ESPECIAS	WAL-MART	Julio 2016	ARG
PP	GATO NEGRO	GATO NEGRO	Enero 2017	ESPAÑA
PP	GATO NEGRO	GATO NEGRO	Enero 2017	HUNGRÍA
PP	DUSEN	MERCADO BAHIA	Enero 2017	ARG
PP	AGROAUCONA	MERCADO BAHIA	Enero 2017	ARG
PDA	GRANJAS PATAGONICAS	LA FUSTA	Enero 2017	ARG
PDA	GATO NEGRO	GATO NEGRO	Enero 2017	ESPAÑA
PPA	GATO NEGRO	GATO NEGRO	Enero 2017	ESPAÑA
PA	DUSEN	MERCADO BAHIA	Enero 2017	ARG
PA	BILL KATHERINO	MERCADO BAHIA	Enero 2017	ARG

PD: Pimentón dulce, PP: Pimentón picante o Paprika, PDA: Pimentón dulce ahumado, PPA: Pimentón picante ahumado, PA: Pimentón ahumado.

Con el fin de lograr una correcta homogenización con un tamaño de partícula uniforme, tanto de las muestras sin adulterar como las adulteradas, se evaluaron diferentes tratamientos para adecuarlas a la técnica espectroscópica utilizada. Se probó la homogenización con morteros de ágata, molinillo eléctrico y tamices de diferentes tamaños. En todos los casos se estableció un tiempo de duración del procesado de 10 minutos. Con el fin de comprobar la efectividad de los distintos tratamientos, las muestras fueron observadas en un estereomicroscopio Olympus trinocular SZ-PT con cámara de video Sony 151A incorporada. En la Figura 3.1 se pueden observar las diferencias entre una muestra original (3.1.a), y procesada con los diferentes tratamientos (3.1.b molinillo, 3.1.c mortero y 3.1.d tamiz), poniéndose en

evidencia que la mejor homogenización se logra mediante el uso del tamiz (ASTM #70).

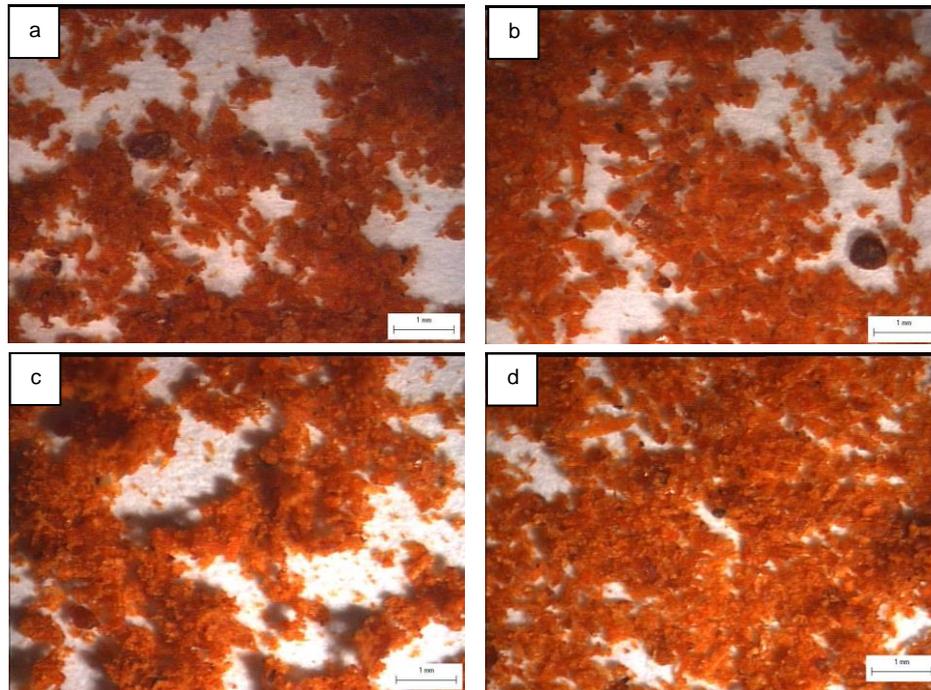


Figura 3.1. Procesados de muestra: (a) muestra original, (b) molinillo eléctrico, (c) mortero de ágata y (d) tamiz ASTM 70 mesh.

La selección del procesado de las muestras originales y adulteradas se comprobó también mediante inspección visual de los espectros NIR de todos los tratamientos, A modo de ejemplo, en la Figura 3.2 se observa una mejor diferenciación espectral mediante la utilización de tamiz con respecto al procesado con molinillo eléctrico.

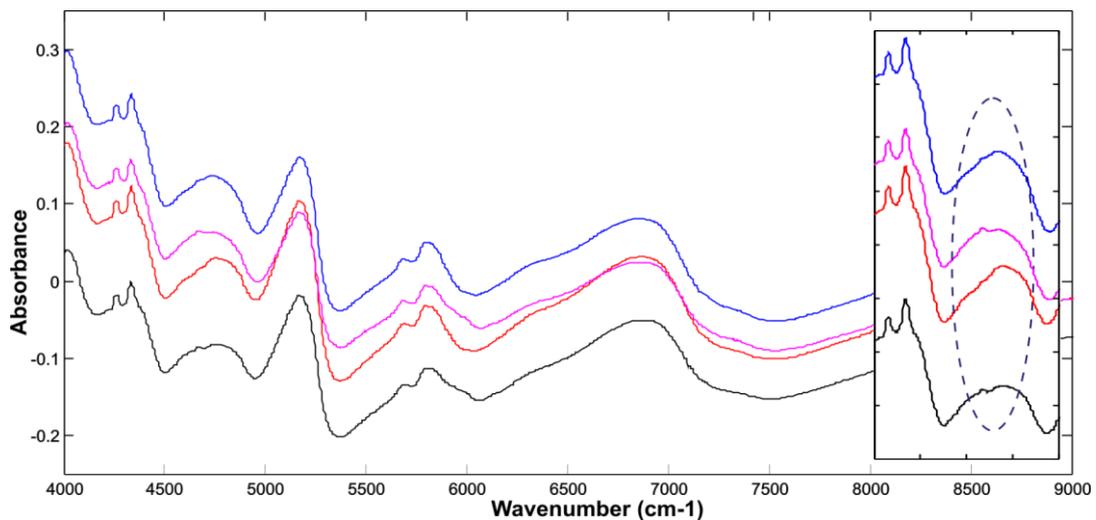


Figura 3.2. Espectros NIR de muestras originales y adulteradas (Sudan I al 5%) procesadas con molinillo eléctrico (azul original, rosa adulterada) y con tamiz (roja original, negra adulterada).

Por consiguiente, en base a lo expuesto anteriormente, las muestras fueron procesadas con un tamiz ASTM #70 tal como se muestra en la Figura 3.3, para lograr la mejor homogenización entre los pimentones y los adulterantes. Una vez tamizadas, las muestras fueron almacenadas en desecador.



Figura 3.3. Tamizado de pimentones.

Los adulterantes utilizados Sudan I, II, III, IV y Para-Red presentan una intensa tonalidad anaranjada-rojiza, tal como se puede apreciar en la Figura 3.4. Todos los colorantes utilizados son de una pureza superior al 95%.



Figura 3.4. Colorantes Sudan I, II, III y IV (arriba de izquierda a derecha) y Para-Red (abajo).

La adulteración de los pimentones se realizó mediante el agregado de los colorantes Sudan I, II, III, IV y Para-Red en concentraciones de 5; 2,5 y 1 %. La Figura 3.5 muestra una fotografía estereoscópica de un pimentón original y uno adulterado

con Sudan I al 5%, observándose una apreciable diferencia de color entre las muestras.

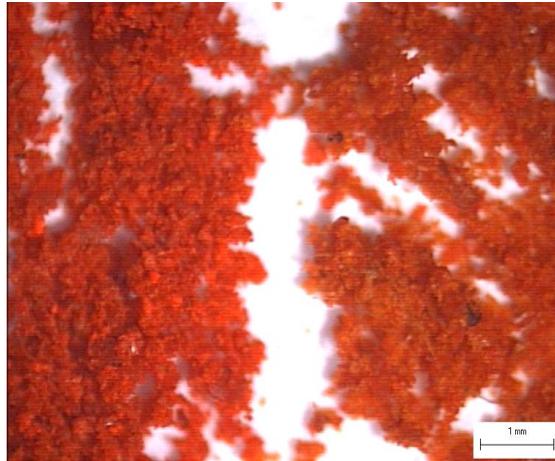


Figura 3.5. Fotografía estereoscópica de una muestra adulterada (izquierda) y una sin adulterar (derecha).

Por otro lado, los pimentones presentaron diferencias tanto en la cantidad como en la calidad de los residuos de tamización, dependiendo del origen (dulce, picante) y del tipo de procesamiento (ahumado) de la matriz (Figura 3.6). Solamente en el caso particular de una de las muestras, más oleosa y húmeda, el procesamiento resultó más dificultoso, correspondiendo al pimentón dulce La Parmesana (PD1).



Figura 3.6. Residuo tamizado muestras.

## 4.2 Obtención de espectros NIR

Las medidas espectroscópicas se realizaron en un instrumento Thermo Scientific Nicolet iS50 FT-IR con módulo NIR. Los parámetros instrumentales optimizados fueron: background, gain y resolución.

Un background es un espectro de referencia que tiene en cuenta las propiedades ópticas del módulo de muestreo y el instrumento, de manera tal que el espectro final queda corregido respecto a la señal del mismo. Se evaluaron dos posibles backgrounds: uno interno que utiliza un material de referencia de oro y uno externo de teflón, el cual posee una alta reflectancia difusa. Como se puede observar en la Figura 3.7, los espectros de los background interno y externo son similares, presentando el interno una mayor absorbancia. Por consiguiente, se decidió utilizar el background externo como referencia.

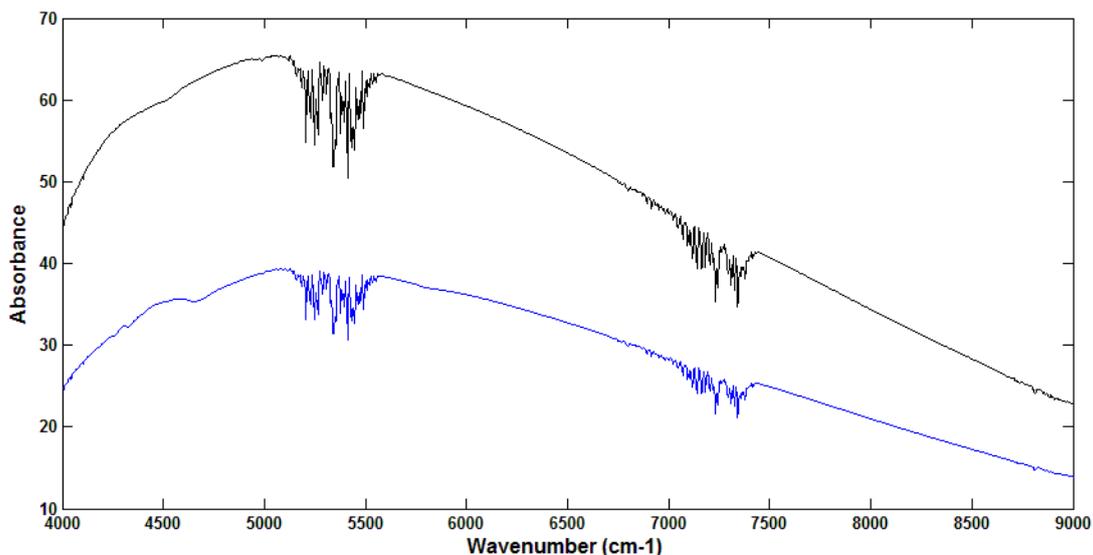


Figura 3.7. Espectros backgrounds interno (negro) y azul (externo).

El parámetro gain amplifica la intensidad de señal del detector, aumentando la relación señal/ruido. Valores grandes de gain son útiles en muestras con señales débiles. Los valores de gain evaluados fueron 1.0, 2.0 y automático (autogain). Por otro lado, para mejorar la relación señal/ruido, se evaluaron valores de resolución de 4, 6, 8 y 16  $\text{cm}^{-1}$ . Los valores óptimos utilizados para la obtención de los espectros fueron una resolución de 8  $\text{cm}^{-1}$  y un valor de gain de 1.0.

Una vez optimizados los parámetros instrumentales, se procedió a obtener los espectros de las 20 muestras originales y las adulteradas con los 5 colorantes (Sudan I, II, III, IV y Para-Red) en los tres niveles de concentración estudiados (5; 2,5 y 1%),

resultando en un total de 320 muestras. Las mismas fueron colocadas en viales de vidrio de 1,8 cm de diámetro, teniendo en cuenta la altura de llenado en cada medición. Se realizaron dos réplicas de cada una, recargando el vial para cada medida. La Figura 3.8 muestra los espectros NIR de los veinte pimentones originales donde se puede observar un perfil espectral similar entre ellos. Sin embargo, existe una muestra que presenta mayor absorbancia correspondiente al pimentón con mayor dificultad en su procesamiento por tener características diferentes, siendo este más oleoso y húmedo (PD1).

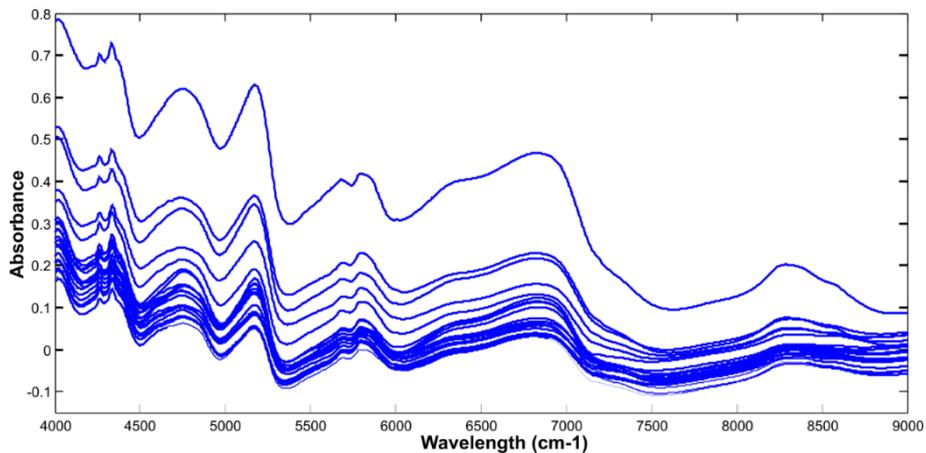


Figura 3.8. Espectros NIR de pimentones originales.

En la Figura 3.9 se observan los espectros NIR de los cinco colorantes puros (Sudan I, II, III, IV y Para-Red) apreciándose que son similares y que las principales señales están comprendidas en las zonas espectrales del 4000 al 5000 y del 5500 al 6500  $\text{cm}^{-1}$ .

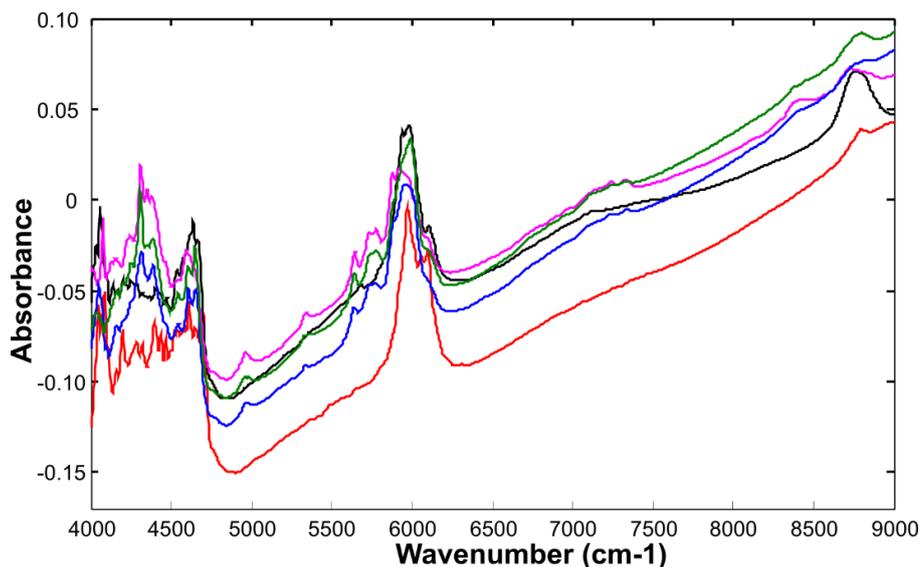


Figura 3.9. Espectros NIR de colorantes Para-Red (rojo), Sudan I (negro), Sudan II (magenta), Sudan III (azul) y sudan IV (verde).

Por otro lado, la Figuras 3.10 y 3.11 muestran los espectros NIR de pimentones seleccionados al azar en sus variedades dulce, picante, ahumado y picante-ahumado, en sus versiones original y adulterados con cada colorante al 5%. En todos los casos, se evidencian diferencias espectrales tanto entre muestras originales y adulteradas como entre las cinco muestras adulteradas. Las zonas donde estas diferencias son más significativas se encuentran comprendidas entre 4500-5000  $\text{cm}^{-1}$  y entre 5400 y 6300  $\text{cm}^{-1}$ .

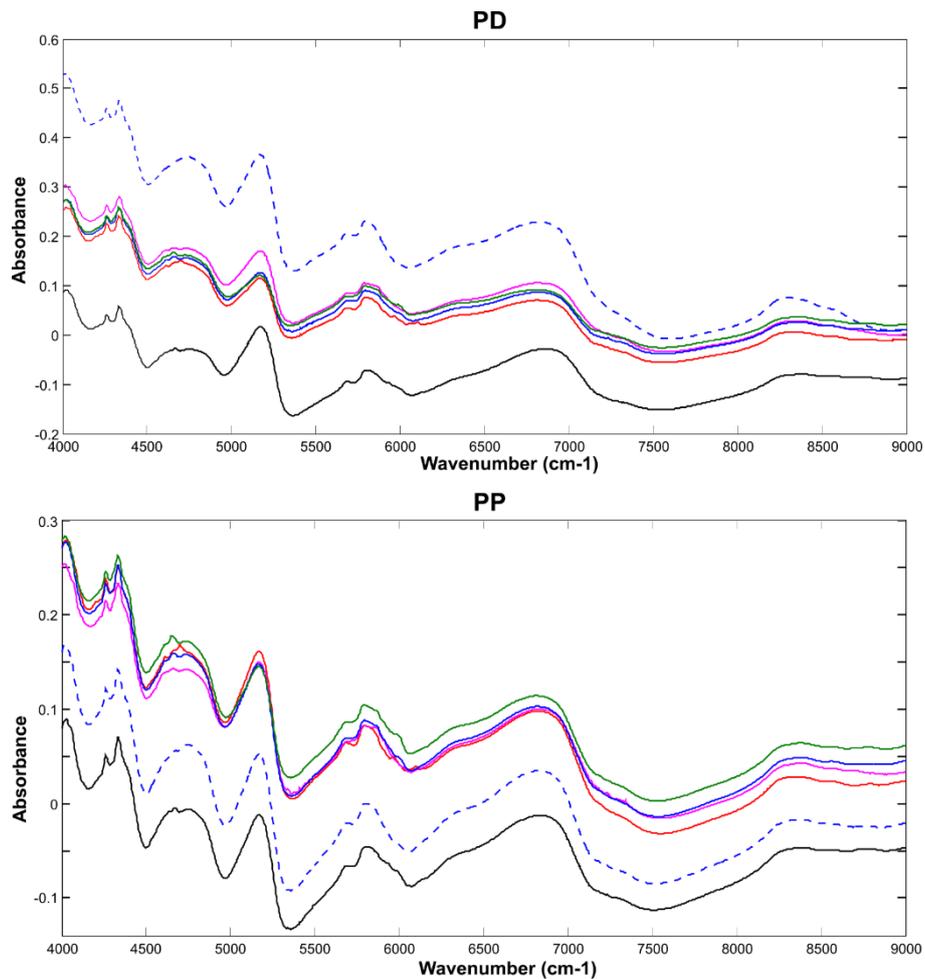


Figura 3.10. Espectros NIR de (a) pimentón dulce, (b) pimentón picante. Muestras sin adulterar (punteado azul) y muestras adulteradas con Para-Red (rojo), Sudan I (negro), Sudan II (magenta), Sudan III (azul) y sudan IV (verde).

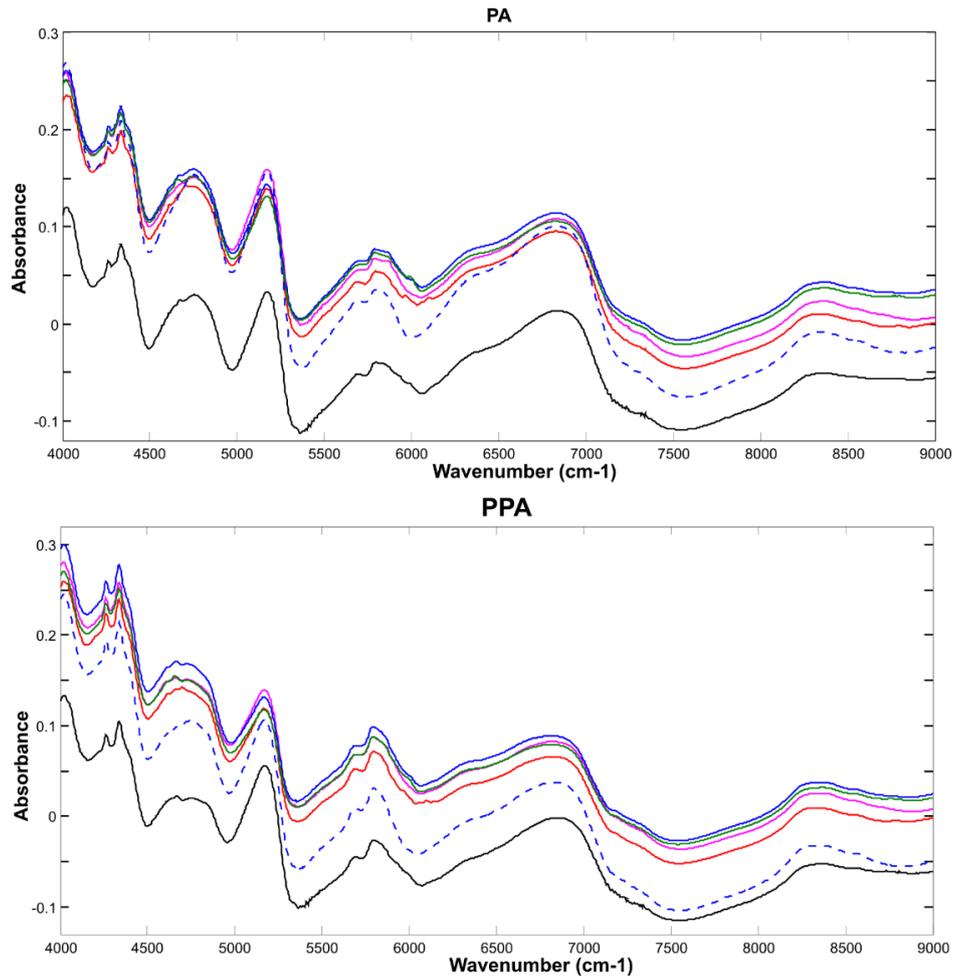


Figura 3.11. Espectros NIR de (a) pimentón ahumado y (b) pimentón picante-ahumado. Muestras sin adulterar (punteado azul) y muestras adulteradas con Para-Red (rojo), Sudan I (negro), Sudan II (magenta), Sudan III (azul) y sudan IV (verde).

### 4.3 Pre-procesamiento de espectros

El pre-procesamiento espectral son correcciones del espectro basadas en herramientas matemáticas para aumentar la calidad de la señal instrumental, reduciendo al máximo la información asociada a la propiedad estudiada. El espectro NIR depende de la composición química de la muestra, de algunas propiedades físicas tales como el tamaño, la forma y la distribución de partículas. Estas características pueden dar lugar a efectos de scattering, asociados a variaciones sistemáticas no deseadas, que complican los modelos multivariantes en detrimento de la calidad de los resultados de los análisis. Esto provoca desplazamientos en la línea de base y efectos de no linealidad, que pueden reducirse mediante la aplicación de pre-tratamientos

adecuados para asegurar que el conjunto de datos incluya todas las posibles fuentes de variabilidad en las muestras. En consecuencia, el preprocesado de espectros es el paso más importante antes de aplicar un modelo multivariate.

Las técnicas de pre-procesamiento más ampliamente utilizadas en NIR, tanto en el modo reflectancia como transmitancia pueden ser divididas en dos categorías: métodos de corrección de scattering y derivadas espectrales. Dentro del primer grupo se utilizaron los métodos MSC (Multiplicative Scatter Correction), SNV (Standard Normal Variate), detrending y wavelets. En el caso de derivadas espectrales, se utilizó el método de Savitsky-Golay para la primera y segunda derivada.

La Figura 3.12 muestra los espectros procesados mediante MSC y SNV. La principal diferencia entre ambos métodos es que el SNV se aplica a cada espectro independientemente, mientras el MSC requiere la definición de un espectro de referencia. En nuestro caso, en MSC se utilizó un espectro promedio como referencia. Se puede observar que en ambos casos la corrección de scattering es similar, aunque con moderadas diferencias en algunas zonas del espectro, por ejemplo, alrededor de 5500  $\text{cm}^{-1}$  y 7500  $\text{cm}^{-1}$ .

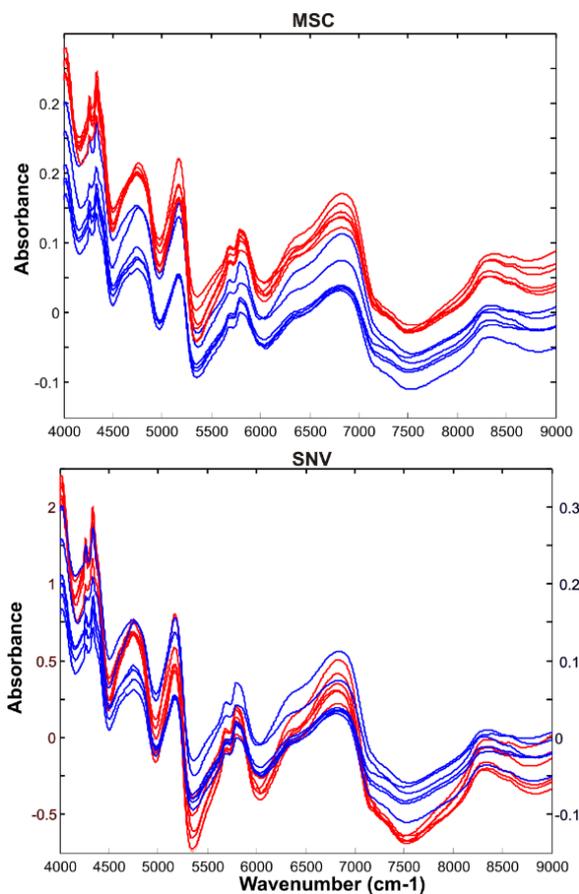


Figura 3.12. Espectros originales (azul) y procesados (rojo) con MSC y SNV.

En la Figura 3.13 se observan los espectros procesados mediante detrending y transformada wavelet, en los que se logra una corrección tanto de scattering como de línea de base. En el caso de detrending, se utilizaron polinomios de grado 1 (rojo) y grado 2 (negro), destacándose la similitud a lo largo de todo el espectro con diferencias a partir de  $8000\text{ cm}^{-1}$ . En el procesamiento mediante wavelets se utilizó una wavelet madre de la familia Daubechies y un nivel de descomposición de 8. Se puede observar que este procesamiento proporciona una línea de base mejor definida y espectros corregidos más uniformes. Esto puede deberse a que a cada espectro NIR se le resta información no relevante asociada al background y la línea de base distorsionada contenidos en el coeficiente de aproximación del nivel seleccionado.

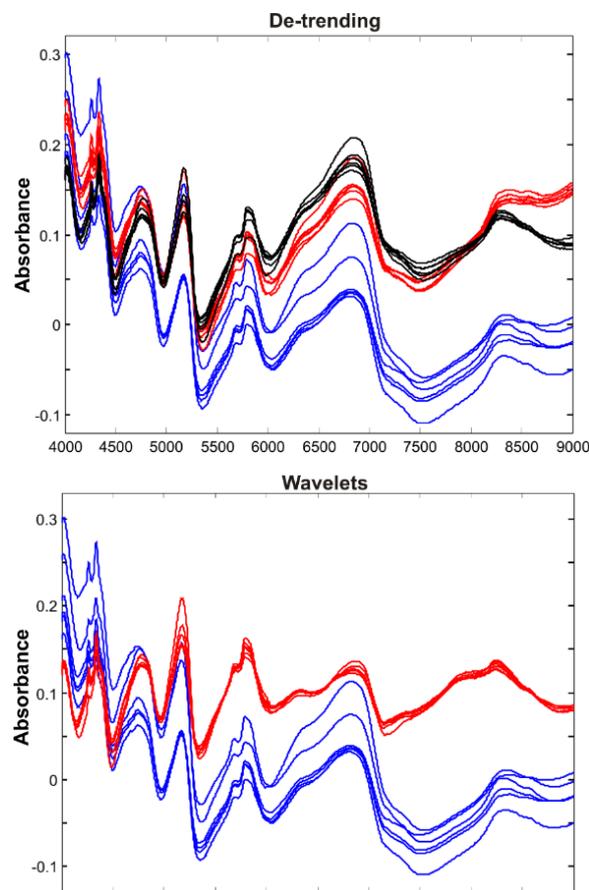


Figura 3.13. De-trending con polinomio de grado 1 (rojo), grado 2 (negro) y Wavelets. Espectros originales (azul) y procesados (rojo).

Finalmente, se pre-procesaron los espectros mediante derivadas. La derivación es muy utilizada en espectroscopia NIR por su capacidad de disminuir los problemas característicos como solapamiento de bandas y variaciones de la línea base. La utilización de la primera derivada elimina desplazamientos constantes de la línea base, mientras que la segunda derivada lo hace con desplazamientos que varían linealmente con el número de onda. En nuestro caso, para el cálculo de las derivadas se utilizó el

método de Savitzky-Golay, con un paso previo de suavizado debido a que la derivada puede amplificar el ruido espectral. Para llevarlo a cabo, es necesario optimizar el tamaño de ventana ya que un incorrecto uso de la misma podría distorsionar la información de los datos originales o resultar en un suavizado insatisfactorio. Por consiguiente, se probaron valores de ventana comprendidos entre 7 y 13, seleccionando como parámetros finales un valor de 11 y un polinomio de grado dos para ambas derivadas. La Figura 3.14 muestra los espectros procesados por ambas derivadas.

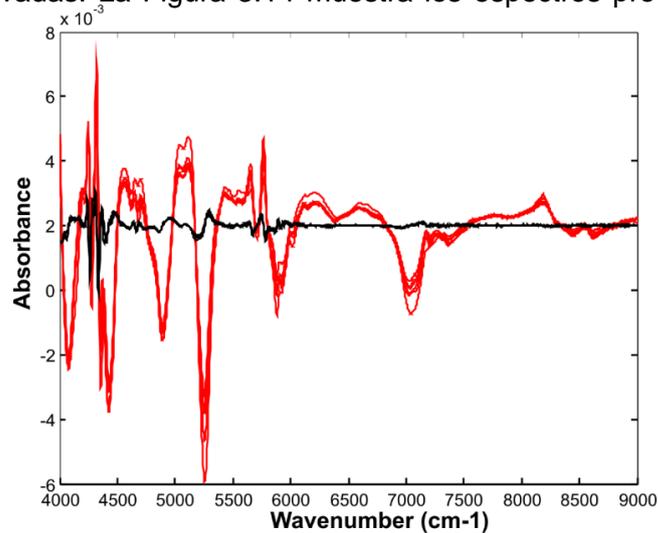


Figura 3.14. Derivadas primera (negro) y segunda (rojo).

#### 4.4 Modelos de clasificación multivariante

Con la finalidad de seleccionar el óptimo pre-procesado, se utilizaron los datos tratados con los siete métodos estudiados y brutos (sin pre-procesar) para construir varios modelos de clasificación multivariante con Partial Least Squares-Discriminant Analysis (PLS-DA). Las clases definidas para los modelos de clasificación son las siguientes (Tabla 3.2):

Tabla 3.2: Codificación de muestras		
Clases	Tipo de muestra	Codificación
Clase 1	Original	1 - 20
Clase 2	Adulterada Para-Red	21 - 40 (5%), 41 - 60 (2,5%), 61 - 80 (1%)
Clase 3	Adulterada Sudan I	81 - 100 (5%), 101 - 120 (2,5%), 121 - 140 (1%)
Clase 4	Adulterada Sudan II	141 - 160 (5%), 161 - 180 (2,5%), 181 - 200 (1%)
Clase 5	Adulterada Sudan III	201 - 220 (5%), 221 - 240 (2,5%), 241 - 260 (1%)
Clase 6	Adulterada Sudan IV	261 - 280 (5%), 281 - 300 (2,5%), 301 - 320 (1%)

La Tabla 3.3 muestra los resultados de clasificación en términos de precisión para cada clase (número de muestras correctamente aceptadas sobre el número total de muestras aceptadas por esa clase). La habilidad de los diferentes modelos de clasificación fue evaluada mediante validación cruzada. (cross-validation) utilizando Venetian blinds (5 grupos de cancelación). El número óptimo de variables latentes (Latent Variables LVs) del modelo PLS-DA se escogió en función del mínimo valor del error por validación cruzada (RMSECV) considerando globalmente las 6 clases.

	<b>C1</b>	<b>C2</b>	<b>C3</b>	<b>C4</b>	<b>C5</b>	<b>C6</b>
<b>Datos brutos</b>	54	91	70	57	44	62
<b>MSC</b>	77	92	81	92	63	82
<b>SNV</b>	71	92	80	95	68	85
<b>Detrend 1</b>	77	89	77	94	69	89
<b>Deetrend 2</b>	80	87	83	79	54	81
<b>Wavelets</b>	69	94	87	81	64	84
<b>Primera derivada</b>	68	92	75	85	76	87
<b>Segunda derivada</b>	<b>80</b>	<b>82</b>	<b>98</b>	<b>97</b>	<b>98</b>	<b>83</b>

Se puede observar que los mejores resultados de clasificación global, es decir considerando todas las clases, corresponden a los datos procesados con la segunda derivada. Por consiguiente, éste fue el seleccionado para continuar con el tratamiento multivariante de los datos.

Posteriormente se realizó un análisis exploratorio con Principal Components Analysis (PCA) en datos originales y procesados con la segunda derivada. La Figura 3.15 muestra el gráfico de scores de los datos originales centrados, con el objeto de eliminar la información común a todas las muestras. El porcentaje de varianza explicado por las tres primeras componentes principales (PCs) es de 99,59%, siendo PC1 la componente que proporciona la mayor cantidad de información (96,39%). En general se puede observar un gran solapamiento entre las distintas clases, apreciándose que la mayoría de muestras de clase 3 (adulteradas con Sudan I, cuadrados azules) tienen valores de scores negativos a lo largo de PC1. Se concluye que la información contenida en esta componente está mayormente asociada a la matriz (pimentón) que a los adulterantes en todas las concentraciones.

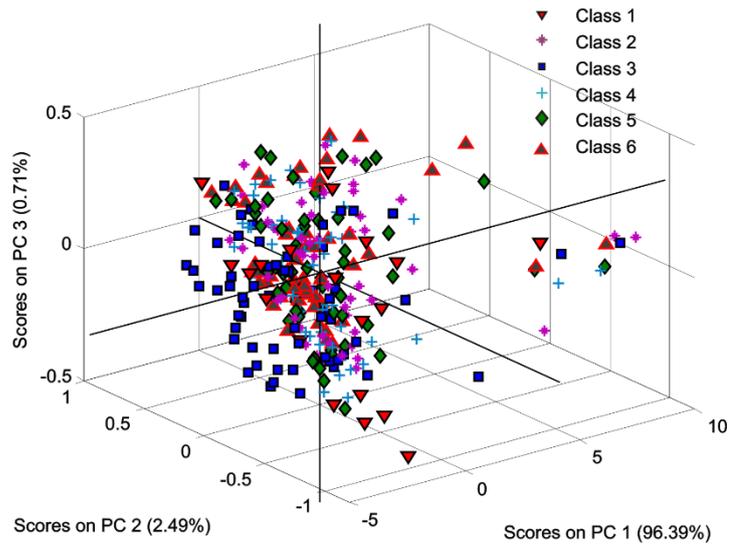


Figura 3.15. Grafico de Scores PCA de datos originales.

Del análisis de PCA de los datos procesados con la segunda derivada (Figura 3.16) con un porcentaje de varianza explicada del 70,58%, se visualiza una discriminación significativamente mejorada entre las seis clases con respecto a los datos sin procesar, con excepción de las clases 1 y 5 que están parcialmente solapadas en el espacio multivariante.

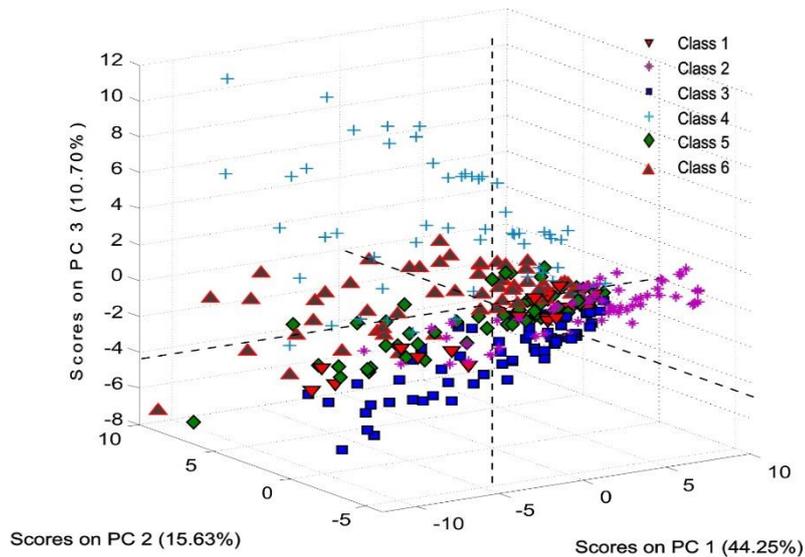


Figura 3.16. Grafico de Scores PCA de datos procesados con la segunda derivada.

#### 4.5 Selección de variables: i-modelos

Con el fin de evaluar una mejora en los resultados de clasificación para lograr una óptima discriminación entre todas las clases, se procedió a realizar una selección de variables en los datos procesados con la segunda derivada, con el fin de retener aquellas con mayor poder de discriminación y eliminar información no redundante. Existe una amplia variedad de algoritmos de selección de variables en literatura, por lo cual, en base a la experiencia previa, se evaluaron métodos basados en intervalos (i-modelos).

Dentro de los i-modelos, se evaluó i-PLS; backward i-PLS; bi-PLS;

- i-PLS: Se dividieron los datos en 20 intervalos, se calcularon modelos PLS en cada uno (con un máximo de 15 LVs) en datos centrados con validación cruzada.
- backward interval PLS (biPLS): Se utilizaron los mismos parámetros que para i-PLS.
- interval PCA (iPCA): Los parámetros usados fueron: 10 PCs como máximo en cada intervalo, datos centrados, 20 intervalos en total.

De todos los algoritmos utilizados en los datos procesados con la segunda derivada, el que proporcionó los mejores resultados de clasificación fue i-PCA. Luego de evaluar los veinte gráficos de scores, los intervalos candidatos a ser seleccionados son los que se muestran en la Tabla 3.4.

<b>Intervalo</b>	<b>Variables involucradas (cm<sup>-1</sup>)</b>
1	4000 – 4247
3	4501 - 4748
8	5755 - 6001
9	6005 - 6252
20	8759 - 9002

A modo de ejemplo, la Figura 3.17 muestra dos de los intervalos seleccionados, en los que se puede observar una satisfactoria discriminación entre las seis clases. Por lo tanto, luego de escoger los cinco intervalos mencionados en la Tabla 3.4, se seleccionaron 324 variables de un total de 1298.

A continuación, en la Figura 3.18, correspondiente al gráfico de scores PCA de los datos reducidos con un porcentaje de varianza explicada de 64,98%, se visualiza

una considerable mejora en la discriminación entre las clases; con respecto al modelo que contiene todas las variables, mostrado en la Figura 3.16. Asimismo, se observa que PC1 contiene información mayoritariamente asociada a las clases 2 (adulteradas con Para-Red) y 3 (adulteradas con Sudan I) mientras que PC3 recoge principalmente la de la clase 4 (adulteradas con Sudan 2).

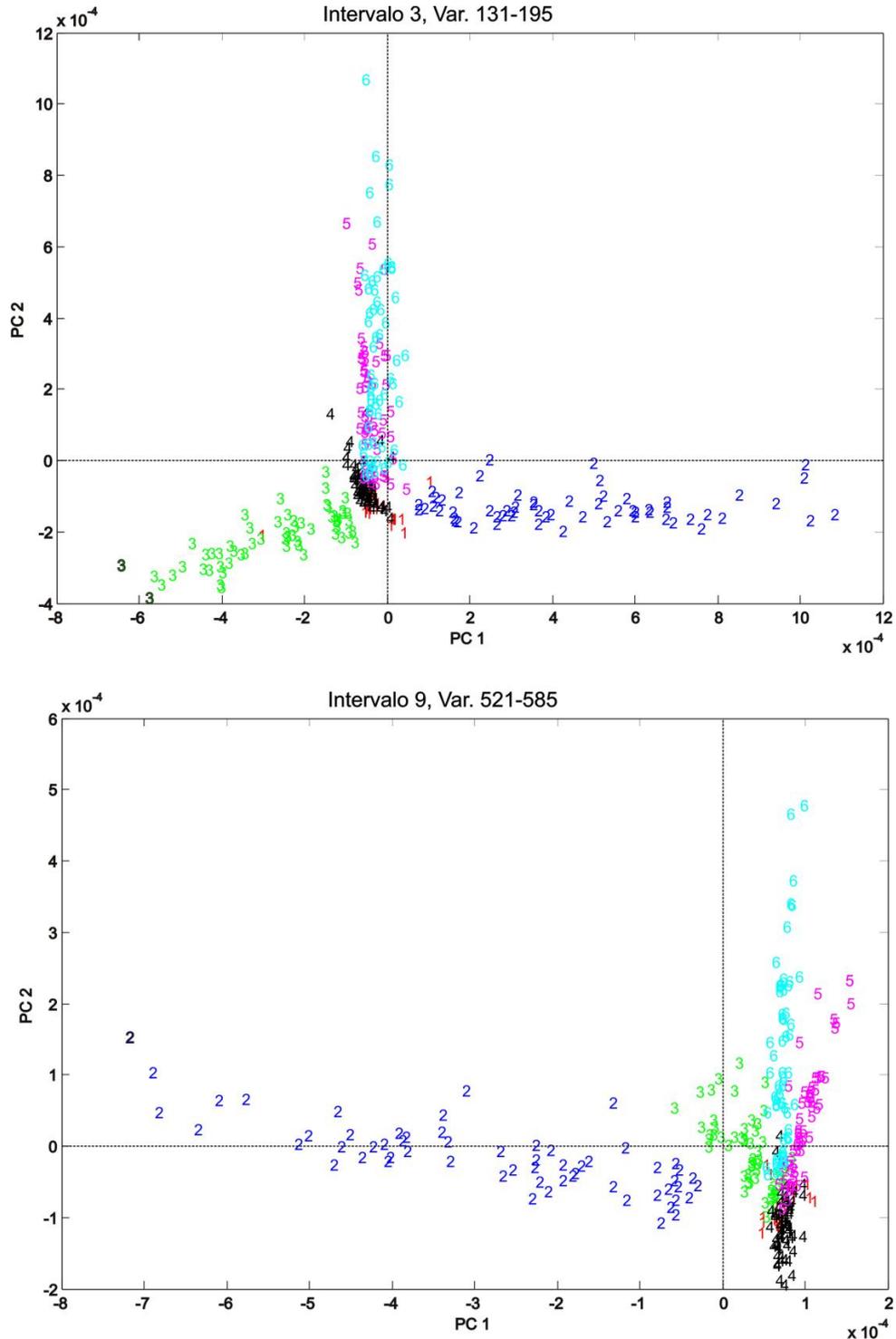


Figura 3.17. Grafico de Scores PCA de los intervalos 3 y 9 seleccionados por i-PCA.

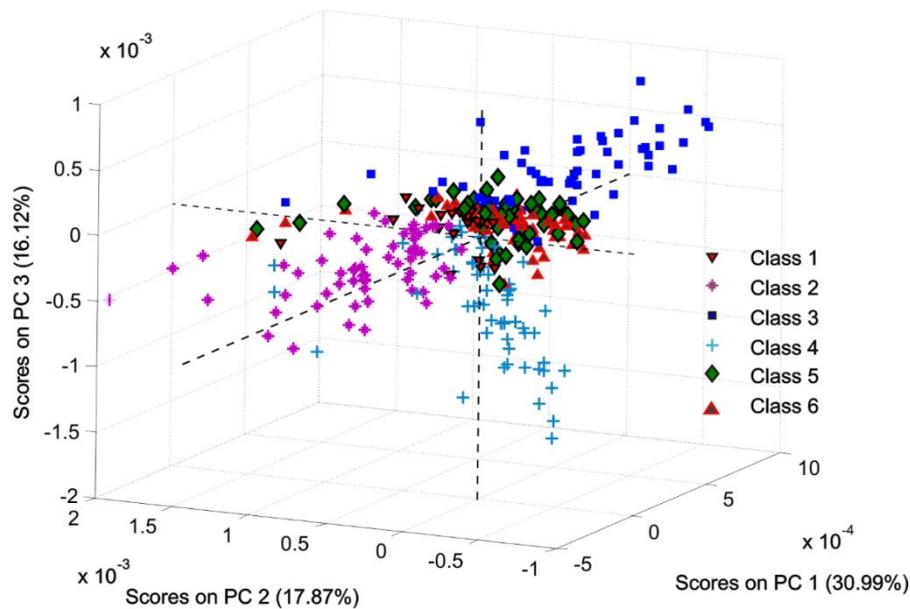


Figura 3.18. Gráfico de Scores PCA de datos procesados con la segunda derivada y con selección de variables i-PCA.

#### 4.6 Evaluación del modelo de clasificación obtenido con i-PCA

Una vez escogidas las variables más relevantes con i-PCA, se construyó un modelo de clasificación con PLS-DA con las 324 variables seleccionadas, evaluando los resultados obtenidos por validación cruzada (Venetian blinds, 5 grupos de cancelación), escogiendo el número óptimo de LVs en función del mínimo RMSECV considerando globalmente todas las clases. La habilidad del modelo de clasificación se evaluó en términos de varios parámetros de calidad, calculados a partir de la obtención de verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN), tal como se muestra en la Tabla 3.3. Cuando se define un espacio para una determinada clase, las muestras se consideran correctamente aceptadas o no, dependiendo si caen dentro o fuera de los límites establecidos para la misma. En particular, los TP representan la fracción de muestras pertenecientes a una clase que son correctamente aceptadas y los TN la fracción de muestras que no pertenecen a la clase y que son correctamente rechazadas. Por otro lado, los FN son muestras pertenecientes a una clase que caen fuera de sus límites y los FP corresponden a muestras de otra/s clases que caen dentro del límite de esa clase.

Tabla 3.3. Posibles resultados de clasificación

Predicciones	Actual	
	Positivo	Negativo
Positivo	<b>TP</b>	<b>FP</b>
Negativo	<b>FN</b>	<b>TN</b>

En nuestro caso particular, considerando la clase 1 (muestras sin adulterar), los verdaderos positivos (TP) son las muestras de la clase correctamente aceptadas y los falsos negativos (FN) corresponden a las muestras sin adulterar asignadas a clases adulteradas. Por otro lado, los verdaderos negativos (TN) son las muestras adulteradas correctamente rechazadas y los falsos positivos (FP) son muestras adulteradas erróneamente asignadas a esta clase. La implicancia de cada uno de estos resultados es muy diferente: la obtención de FN conlleva un riesgo económico, ya que muestras sin adulterar serían innecesariamente retiradas de su comercialización, debiendo ser sometidas a un análisis confirmatorio. La importancia de los FP es que representan un riesgo para la salud del consumidor, ya que muestras adulteradas serían aceptadas como muestras aptas para el consumo, pudiendo provocar un grave impacto en la salud pública.

La Tabla 3.4 muestra los parámetros de calidad del modelo de clasificación obtenido con selección de variables i-PCA. La sensibilidad representa una medida del nivel de confianza de la clase, la especificidad es la capacidad de reconocer muestras extrañas y se relaciona con la obtención de falsos positivos, la eficiencia del modelo da una idea de la habilidad global de clasificación y puede variar entre 0 cuando tanto la sensibilidad como la especificidad tienen valores nulos y 1 cuando ambos parámetros son máximos. La precisión se interpreta como la probabilidad de que un resultado positivo sea realmente correcto, el índice de Youden representa el punto de corte que determina la sensibilidad y especificidad más alta evaluadas conjuntamente y finalmente el coeficiente de correlación de Matthews (MCC) considera los cuatro posibles resultados, pudiendo variar entre valores de +1 (todas las muestras perfectamente asignadas), -1 (todas las muestras incorrectamente asignadas) y 0 para asignaciones azarosas.

**Tabla 3.4. Parámetros de calidad de un modelo de clasificación**

$$\text{Sensibilidad} = TP / (TP + FN)$$

$$\text{Especificidad} = TN / (TN + FP)$$

$$\text{Eficiencia} = (TN + TP) / (TN + FP + TP + FN)$$

$$\text{Precisión} = TP / (TP + FP)$$

$$\text{Índice Youden} = (\text{sensibilidad} + \text{especificidad} - 1)$$

$$\text{Coeficiente correlación Matthews} = \frac{(TP \times TN) - (FP \times FN)}{[(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)]^{1/2}}$$

La Tabla 3.5 muestra los resultados obtenidos; donde se puede aseverar globalmente que todos los parámetros de calidad son muy satisfactorios. Considerando la clase 1, cabe resaltar la alta especificidad obtenida, evidenciando una muy buena capacidad de reconocer muestras adulteradas a diferentes niveles de concentración (1, 2,5 y 5%), lo cual es muy positivo desde el punto de vista de la salud del consumidor. Por otro lado, la sensibilidad es superior al 95%, por lo que el riesgo económico asociado a la obtención de falsos negativos es bajo. Considerando los otros parámetros de calidad, el modelo presenta una excelente eficiencia y precisión (99 y 95%, respectivamente), lo que representa una alta probabilidad de que una decisión positiva sea correcta. También se obtuvo un índice de Youden cercano a 100 y coeficiente de correlación de Matthews muy cercano a +1; demostrando una alta capacidad de asignación correcta de muestras. Cabe aclarar que, considerando las muestras adulteradas, los pocos errores de clasificación obtenidos provienen de muestras con Sudan IV al 1%, como era de esperar del nivel más bajo de concentración. Por otro lado, en relación a las cinco clases adulteradas, se obtuvieron excelentes resultados considerando globalmente todos los parámetros. En base a todo lo anteriormente expuesto, se puede concluir que el modelo de clasificación es altamente confiable.

**Tabla 3.5: Resultados de clasificación**

	C1	C2	C3	C4	C5	C6
<b>Sensibilidad</b>	95,0	100	100	100	100	96,7
<b>Especificidad</b>	99,7	100	99,6	100	100	100
<b>Eficiencia</b>	99,4	100	99,7	100	100	99,4
<b>Precisión</b>	95,0	100	98,4	100	100	100
<b>Youden's index</b>	94,7	100	98,1	100	100	96,7
<b>MCC</b>	0,95	1,00	0,99	1,00	1,00	0,98

Finalmente se evaluaron las curvas ROC, herramienta visual que permite evaluar la eficiencia del modelo de una manera muy útil, resumiendo la relación entre sensibilidad y especificidad. Estas curvas se obtienen de graficar los valores de sensibilidad contra  $1 - \text{especificidad}$  (o tasa falsos positivos) para cada parámetro que define los límites entre las clases de las técnicas de clasificación. En el caso particular de PLS-DA, se grafica para el valor umbral de cada clase obtenido en base al teorema de Bayes (Fawcett T., 2006) hasta encontrar el valor que proporcione la óptima sensibilidad y especificidad. La Figura 3.19 muestra, a modo de ejemplo, las curvas ROC obtenidas para las clases 1 y 5. Como se puede observar, las curvas arrancan en (0,0) que corresponde a una máxima especificidad y a una nula sensibilidad, y terminan en (1,1) donde la sensibilidad es máxima y la especificidad nula. El valor óptimo es (0,1), que corresponde al cruce entre ambas rectas (100% sensibilidad y especificidad), este comportamiento se puede evidenciar en todas las clases consideradas.

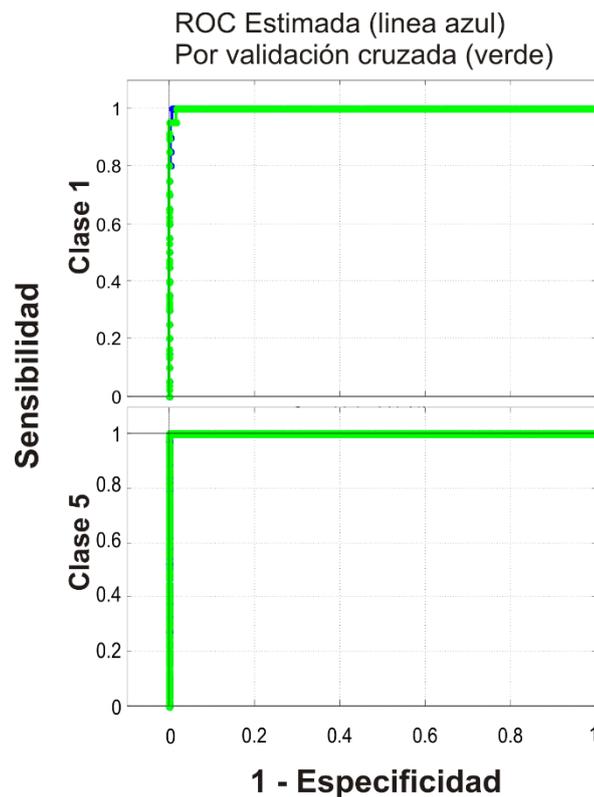


Figura 3.19. Curvas ROC de clases 1 y 5.

Por consiguiente, el modelo de clasificación desarrollado presenta una perfecta capacidad de distinguir falsos positivos de verdaderos positivos, por lo que, en el caso

de adulteración estudiado, son resultados muy satisfactorios considerando la implicancia que estos errores de clasificación tienen en la salud del consumidor.

#### 4.7 Evaluación de la capacidad predictiva del modelo

Una vez obtenidos los resultados de clasificación por validación cruzada, se realizó la evaluación de la capacidad predictiva del modelo de clasificación previamente desarrollado a partir de un conjunto de datos de prueba (test set) conformado por muestras adulteradas. Para ello, se seleccionaron aleatoriamente 18 muestras de las cinco clases adulteradas (6 de cada nivel de concentración) cuya distribución se muestra en la Tabla 3.6.

**Tabla 3.6: Selección de muestras para datos de prueba**

Para-Red			Sudan I			Sudan II			Sudan III			Sudan IV		
5%	2,5%	1%	5%	2,5%	1%	5%	2,5%	1%	5%	2,5%	1%	5%	2,5%	1%
21	41	61	81	101	121	141	161	181	201	221	241	261	281	301
22	42	62	82	102	122	142	162	182	202	222	242	262	282	302
23	43	63	83	103	123	143	163	183	203	223	243	263	283	303
24	44	64	84	104	124	144	164	184	204	224	244	264	284	304
25	45	65	85	105	125	145	165	185	205	225	245	265	285	305
26	46	66	86	106	126	146	166	186	206	226	246	266	286	306
27	47	67	87	107	127	147	167	187	207	227	247	267	287	307
28	48	68	88	108	128	148	168	188	208	228	248	268	288	308
29	49	69	89	109	129	149	169	189	209	229	249	269	289	309
30	50	70	90	110	130	150	170	190	210	230	250	270	290	310
31	51	71	91	111	131	151	171	191	211	231	251	271	291	311
32	52	72	92	112	132	152	172	192	212	232	252	272	292	312
33	53	73	93	113	133	153	173	193	213	233	253	273	293	313
34	54	74	94	114	134	154	174	194	214	234	254	274	294	314
35	55	75	95	115	135	155	175	195	215	235	255	275	295	315
36	56	76	96	116	136	156	176	196	216	236	256	276	296	316
37	57	77	97	117	137	157	177	197	217	237	257	277	297	317
38	58	78	98	118	138	158	178	198	218	238	258	278	298	318
39	59	79	99	119	139	159	179	199	219	239	259	279	299	319
40	60	80	100	120	140	160	180	200	220	240	260	280	300	320

Los resultados de predicción obtenidos se muestran en la Tabla 3.7. Como se puede observar, las muestras adulteradas con los cinco colorantes en sus tres niveles de concentración fueron perfectamente predichas, evidenciando un 100% tanto de sensibilidad como especificidad para todas las clases involucradas. Cabe resaltar la

excelente capacidad predictiva del modelo considerando la ausencia de falsos positivos, ya que estos son los que afectarían la salud del consumidor por ser muestras con colorantes peligrosos asignadas como no adulteradas.

**Tabla 3.7: Resultados de predicción de muestras adulteradas**

	<b>C1</b>	<b>C2</b>	<b>C3</b>	<b>C4</b>	<b>C5</b>	<b>C6</b>
<b>C2</b>	0	18	0	0	0	0
<b>C3</b>	0	0	18	0	0	0
<b>C4</b>	0	0	0	18	0	0
<b>C5</b>	0	0	0	0	18	0
<b>C6</b>	0	0	0	0	0	18

#### 4.8 Selección de variables con SPA-LDA

Paralelamente se evaluó otra técnica de selección de variables, de diferente naturaleza: Algoritmo de Proyecciones Sucesivas y Análisis Discriminante (SPA-LDA). Al conjunto de datos se lo dividió en datos de entrenamiento y prueba (70/30%) mediante el algoritmo de Kennard-Stone, y en la Tabla 3.8 se muestran los resultados de clasificación obtenidos por LDA, con diferentes pre-procesamientos (offset, SNV y MSC). Se puede observar excelentes resultados obtenidos en términos de tasa de clasificación correcta, sensibilidad y especificidad para todas las clases. Cabe destacar que el procesado con corrección offset proporcionó un 100% de tasa de clasificación correcta, sensibilidad y especificidad en todas las clases estudiadas, tanto en datos de entrenamiento (training) como de prueba (test).

Tabla 3.8: Resultados de clasificación con SPA-LDA

		SPA-LDA											
		Training						Test					
%		C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6
	TCC	100						100					
Offset	Sensibilidad	100	100	100	100	100	100	100	100	100	100	100	100
	Especificidad	100	100	100	100	100	100	100	100	100	100	100	100
	TCC	98,2						90,6					
SNV	Sensibilidad	92,9	100	95,2	100	100	97,6	100	100	72,2	88,9	100	88,9
	Especificidad	99,0	100	100	100	99,4	99,4	96,7	100	100	100	93,6	98,7
	TCC	98,2						90,6					
MSC	Sensibilidad	92,9	100	95,2	100	100	97,6	100	100	72,2	88,9	100	88,9
	Especificidad	99,0	100	100	100	99,4	99,4	96,7	100	100	100	93,6	98,7

TCC: Tasa de clasificación correcta

La Figura 3.20 muestra las variables seleccionadas tanto con i-PCA (intervalos grises) como con SPA-LDA (↓ ↓), pudiéndose observar que algunas de ellas resultan seleccionadas por ambas técnicas (↓ ↓). La diferencia en los resultados obtenidos se debe, en parte, a que SPA-LDA selecciona variables puntuales mientras que los i-modelos seleccionan intervalos de variables que pueden contener variables discriminantes como también información no relevante.

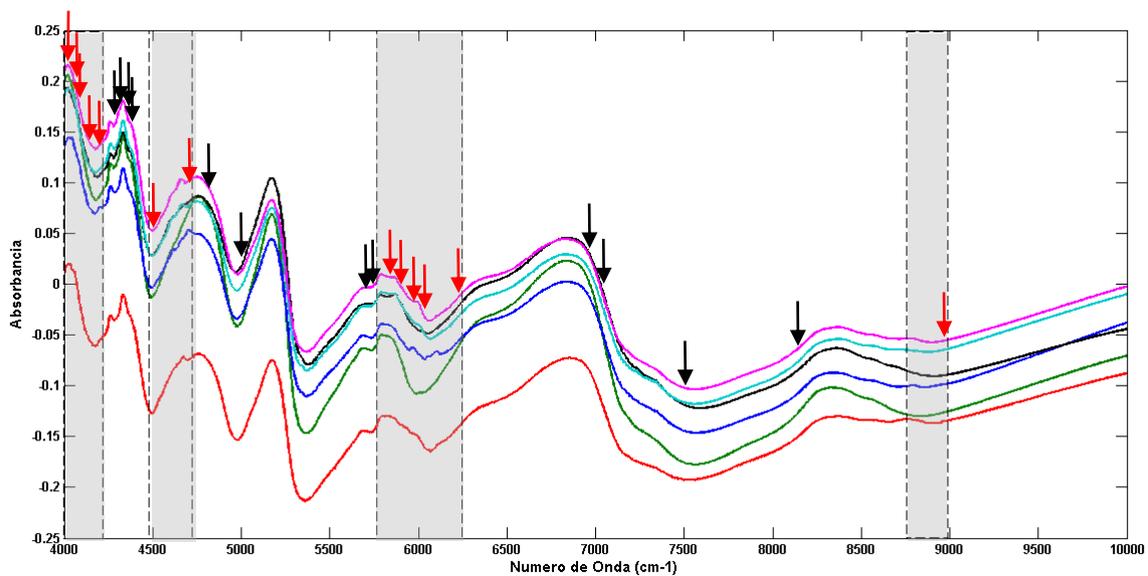


Figura 3.20. Variables seleccionadas con i-PCA (rectángulos) y SPA-LDA (flechas).

Actualmente, a pesar de las regulaciones en materia de seguridad alimentaria, es posible encontrar adulteraciones por el uso de colorantes no permitidos en especias culinarias, por lo tanto, es de vital importancia su control de calidad. En este contexto, el desarrollo de metodologías analíticas de screening son de gran utilidad como herramientas de monitoreo en puntos críticos y de vulnerabilidad en la cadena de suministro de alimentos, con la finalidad de disminuir riesgos en la salud pública asociados con la presencia de sustancias altamente nocivas. La metodología analítica desarrollada en esta tesis representa una herramienta rápida, asequible y confiable, para determinar la adulteración de pimentones con colorantes Sudan y Para-Red, con el potencial de poder ser aplicada para análisis in-situ mediante equipos portátiles.

## 5. CONCLUSIONES

Durante el transcurso de esta tesis ha sido posible desarrollar un método analítico de screening que permite la determinación de la adulteración de pimentones con colorantes no permitidos (Sudan I, II, III, IV y Para-Red), resultando relevante por su grave impacto en la salud pública.

- El tratamiento aplicado a las muestras es simple, sólo se someten a tamización para lograr una correcta homogenización, sin la necesidad de aplicación de procesos de extracción y pre-tratamientos complejos.
- La técnica espectroscópica NIR reduce considerablemente los costos de análisis en comparación con los métodos comúnmente utilizados para determinar colorantes no permitidos en especias culinarias. Esto se debe a la rapidez de las mediciones, instrumentación menos costosa y la no generación de residuos químicos, con potencial aplicación para análisis *in-situ*.
- En relación al análisis multivariante se concluye:
  - De todos los pre-procesamientos espectrales evaluados, la segunda derivada resulta ser el más adecuada en términos de los resultados de clasificación obtenidos, permitiendo una mejor discriminación entre las bandas de los diferentes adulterantes y las muestras originales.
  - El análisis exploratorio de datos (PCA) permite una inspección visual del comportamiento de los datos a través de la agrupación natural de los mismos, posibilitando un mejor entendimiento de los mismos.
- La construcción de un modelo de clasificación con PLS-DA permite predecir si una muestra está adulterada a un determinado nivel de concentración, y que tipo de adulterante contiene.
- La aplicación de estrategias como selección de variables para mejorar resultados de clasificación resultó altamente satisfactoria para tal fin, principalmente en el caso de errores de clasificación con implicancia en la salud del consumidor, los cuales pudieron ser minimizados o eliminados.
- La metodología desarrollada fue satisfactoriamente validada mediante el estudio de parámetros de calidad del modelo de clasificación.

La metodología analítica desarrollada, basada en NIR y análisis multivariante, representa una herramienta de screening fiable, rápida, simple y asequible para determinar adulteraciones alimentarias de gran impacto en la sociedad como lo es la ingesta de colorantes de reconocida peligrosidad. Cabe resaltar que la misma podría

ser implementada en laboratorios de control de rutina de alimentos y gubernamentales. En este sentido, se está trabajando en la elaboración de un dispositivo portátil integrado para análisis *in-situ* en colaboración con el el Laboratorio de Ciencias de las Imágenes (LCI) del Instituto de Investigaciones en Ingeniería Eléctrica de la UNS, con el fin de poder detectar los colorantes de la familia Sudan de forma rápida y económica en los puntos de interés.

## 6. BIBLIOGRAFÍA

Ammu Rosin J., Unni S., Shalini M. y Krishnapillai G. (2016). *Analytical Methods* 8, 5701-5706.

An Y., Jiang L., Cao J., Geng C., y Zhong L (2007). *Mutation Research* 627, 164-170.

Araújo Vilar D., Araujo Vilar M., Lima e Moura T., Nervo Raffin F., Oliveira M., Oliveira Franco C., Filgueiras de Athayde P., Formiga Melo Diniz F., y Barbosa J. (2014). *The Scientific World Journal - Volume 2014*, 11 pag.

Benmassaouda Y., Villaseñor M. J., Salghic R., Jodehd S., Algarrae M., Zougaghb M., y Ríos A. (2017). *Talanta* 166, 63-69.

Berlina A., Zherdev A., Xu Ch., Eremin S., y Dzantiev B. (2017). *Food Control* 73, 247-253.

Brereton R. (2009). *Chemometrics for Pattern Recognition*. Chichester, UK: John Wiley and Sons.

Carreiro Soares S., Gomes A., Rodrigues Galvao Filho A., Ugulino Araujo M., y Kawakami Harrop Galvao R. (2013). *Trends in Analytical Chemistry* 42, 84-98.

CBI Market Intelligence. (2015). *CBI trade Statistics: Spices and herbs*. The Hague, The Netherlands: CBI.

Cozzolino D. (2015). *Analytical methods* 7, 9390-9400.

Chailapakul O., Wonsawat W., Siangproh W., Grudpan K., Zhao Y., y Zhu Z. (2008). *Food Chemistry* 109, 876-882.

Chen D., Li X., Tao Y., Pan Y., Wu Q., Liu Z., Peng D., Wang X., Huang L., Wang Y., y Yuan Z. (2013). *Journal of Chromatography B* 939, 45-50.

Chocholouš P., Gil R., Acebal C., Kubala V., Šatínský D., y Solich P. (2017). *Journal of Separation Science* 40, 1225-1233.

Código Alimentario Argentino, Cap. XVIII, Aditivos Alimentarios. [http://www.anmat.gov.ar/alimentos/codigoa/CAPITULO\\_XVIII.pdf](http://www.anmat.gov.ar/alimentos/codigoa/CAPITULO_XVIII.pdf).

Commision Decision 2003/460/EC. Official Journal of the European Union L154/114.

Commision Decision 2004/92/EC. Official Journal of the European Union L27/52.

Commision Decision 2005/402/EC. Official Journal of the European Union L135/34.

Commision Decision 2009/669/EC. Official Journal of the European Union L194/11.

Di Anibal C., Ruisanchez I., y Callao M. (2011). *Food Chemistry* 124, 1139-1145.

Di Anibal C., Marsal L., Callao M., y Ruisanchez I. (2012). *Spectrochimica Acta Part A* 87, 135-141.

Di Anibal C., Rodríguez M.S., y Albertengo L. (2014). *Food Analytical Methods* 7, 1090-1096.

Di Anibal C., Rodríguez M.S. y Albertengo L. (2015). *Food Control* 56, 18-23.

Di Anibal C., Rodríguez S., Rodríguez M.S., y Albertengo L. (2016). *Food Analytical Methods* 9, 3117-3124.

Du F., Zheng X., Deng J., Zou J., Zeng Q., Li J., y Ruan G. (2017). *Food Analytical Methods* 10, 2018-2026.

Ellis D., Muhamadali H., Simon A., Haughey S., Christopher T., Elliott C., y Goodacre R. (2015). *Analytical Methods* 7, 9401-9414.

Ensafi A., Rezaei B., Amini M., y Heydari-Bafrooei E. (2012). *Talanta* 88, 244-251.

Fawcett T. (2006). *Pattern Recognition Letters* 27, 861-874.

Fonovich T. (2013). *Drug and Chemical Toxicology* 36, 343-352.

Ge Y., Morgan C., Thomasson J., y Waiser T. (2007). *Transactions of the ASABE (American Society of Agricultural and Biological Engineers)* 50, 303-311.

Gómez M., Arancibia V., Aliaga M., Núñez C., y Rojas-Romo C. (2016) *Food Chemistry* 212, 807-813.

Hernández-Hierro J., Garcia-Villanova R. y González-Martín I. (2008). *Analytica Chimica Acta*, 622, 189–194.

Haughey S., Galvin-King P., Ho Y., Bell S., y Elliott C. (2015). *Food control* 48, 75-83.

Horwitz, W. (2002). *Spices and Other Condiments, Color Extractable in Spices*. En: Official Methods of Analysis of the Association of Official Analytical Chemists International. 17th ed (V.II). Gaithersburg: AOAC.

Hu M., Wu L., Song Y., Li Z., Ma Q., Zhang H., y Wang Z. (2016). *Food Analytical Methods*, 9 issue 7, 2124-2132.

IARC (International Agency for Research on Cancer) (1975). *IARC monographs on the evaluation of the carcinogenic risk of chemicals to man: some aromatic azo compounds* 8:224-231. Lyon: Francia.

Jahn M., Patze S., Bocklitz T., Weber K., Cialla-May D., y Popp J. (2015). *Analytica Chimica Acta* 860, 43-50.

Ji W., Zhang M., Wang T., Wang X., Zheng Z., y Gong J. (2017). *Talanta* 165, 18-26.

Karimi S., Feizy J., Mehrro F., y Farrokhnia M. (2016). *RSC Advances* 6, 23085-23093.

Li B., Luo J., Luo H., y Li N. (2015). *Food Chemistry* 173, 594-599.

Li J., Ding X., Liu D., Guo F., Chen Y., Zhang Y., y Liu H. (2013). *Journal of Chromatography B* 942-943, 46-52.

Li W., Zhang R., Wang H., Jiang W., Wang L., Li H., Wu T., y Du Y. (2016). *Analytical methods* 8, 2887-2894.

Liu J., y Gong Z. (2012). *Chinese Journal of Chromatography* 30, 624-629.

Lohumi S., Joshi R., Kandpal L., Lee H., Kim M., Cho H., Mo Ch., Seo Y., Rahman A., y Cho B. (2017). *Food Additives & Contaminants: Part A*, 1-9.

Lohumi S., Lee S., Lee H., y Cho B-K. (2015). *Trends in Food Science & Technology* 46, 85-98.

Marini F. (2013). *Chemometrics in food analysis*. Netherlands: Elsevier.

Mejia E., Ding Y., Mora M.F., y Garcia C. (2007). *Food Chemistry* 102, 1027-1033.

Mercader M.B. (2009). *Avances en la aplicación de la espectroscopía NIR en la industria farmacéutica. Introducción a PAT y técnicas de imagen*. Tesis Doctoral, Universitat Autònoma de Barcelona.

MERCOSUR. (2008). Lista de Sustancias Colorantes Permitidas para Productos de Higiene Personal, Cosméticos y Perfumes. Mercosur/XXXI SGT N° 11/P. RES N° 14/08.

Mishra K., Dixit S., Purshottam S., Pandey R., Das M., y Khanna S. (2007). *International Journal of Food Science & Technology* 42, 1363-1366.

Mojdeh E., Mohammad A., y Hassan K. (2013). *Food Chemistry* 141, 4311-4317.

Moore J., Spink J., y Lipp M. (2012). *Journal of Food Science* 77-4, 118-126.

Muñoz-Olivas R. (2004). *Trends in Analytical Chemistry* 23, 203-216.

Oliveri P. y Downey G. (2012). *Trends in Analytical Chemistry* 35, 74-86.

Otto M. (2017). *Chemometrics: Statistics and Computer Application in Analytical Chemistry*. Germany: Wiley-VCH Verlag GmbH & Co.

Ou Y., Pei L., Lai K., Huang Y., Rasco B., Wang X., y Fan Y. (2017). *Food Analytical Methods* 10, 565-574.

Palanisamy S., Thangavelu K., Chen S., Velusamy V., y Ramaraj S. (2017). *Journal of Electroanalytical Chemistry* 794, 64-70.

Petrakis Eleftherios A., Cagliani L., Tarantilis P., Polissiou M., y Consonni R. (2017). *Food Chemistry* 217, 418-424.

RASFF Reporte Anual 2012:

[https://ec.europa.eu/food/sites/food/files/safety/docs/rasff\\_annual\\_report\\_2012\\_en.pdf](https://ec.europa.eu/food/sites/food/files/safety/docs/rasff_annual_report_2012_en.pdf)

Rebane R., Leito I., Yurchenko S., y Herodes K. (2010). *Journal of Chromatography A* 1217, 2747-2757.

Ren G., y Chen F. (1999). *Journal of Agricultural and Food Chemistry* 47, 2771-2775.

Reinholds I., Bartkevics V., Silvis I., van Ruth S., y Esslinger S. (2015). *Journal of Food Composition and Analysis* 44, 56-72.

Rinnan A., van den Berg F., y Engelsen S. (2009). *Trends in analytical chemistry* 28, 1202-1222.

Schummer C., Sassel J., Bonenberger P., y Moris G. (2013). *Journal of Agricultural and Food Chemistry* 61, 2284-2289.

- Schulz H., Baranska M. (2007). *Vibrational Spectroscopy* 43, 13–25.
- Sciuto S., Esposito S., Dell'Atti L., Guglielmetti C., Acutis P., y Martucci F. (2017). *Journal of Food Protection* 80, 640-644.
- Song Y., Wu L., Li N., Hu M., y Wang Z. (2015). *Talanta* 135, 163-169.
- Tanaka K., Kuba Y., Sasaki T., Hiwatashi F., y Komatsu K. (2008). *Journal of Agricultural Food Chemistry* 56, 8787–8792.
- Tripathi M., Khanna S., y Das M. (2007). *Food Control* 18, 211-219.
- Vandeguiste B.G.M., Massart D.L., Buydens L.M.C., De Jong S., Lewi P.J., y Smeyers-Verbeke J. (1998). *Supervised pattern recognition*. En: Handbook of chemometrics and qualimetrics: part B. Netherlands: Elsevier.
- Wierna R.; Davies L.; Orell R., Colodro V; Nazareno M., y Bonini N. (2015). *La Alimentación Latinoamericana* 315, 62-69.
- Xiaobo Z., Jiewena Z., Povey M. J. W., Holmes M., y Hanpin M. (2010). *Analytica Chimica Acta* 667, 14–32.
- Xu H., Heinze T., Chen S., Cerniglia C., y Chen H. (2007). *Applied and Environmental Microbiology* 73, 7759-7764.
- Yu W., Liu Z., Li Q., Zhang H., y Yu Y. (2015). *Food Chemistry* 173, 815-820.
- Zalacain A., Ordoudi S., Daz-Plaza E., Carmona M., Blazquez I., Tsimidou M. y Alonso G. (2005). *Journal of Agricultural and Food Chemistry* 53, 9337-9341.
- Zhou M., Chen X., Xu Y., Qu J., Jiao L., Zhang H., Chen H., y Chen X. (2013). *Dyes and Pigments* 99, 120-126.
- Zvereva E., Zaichik B., Eremin S., Zherdev A., y Dzantiev B. (2016). *Journal of Analytical Chemistry* 71, 944-948.