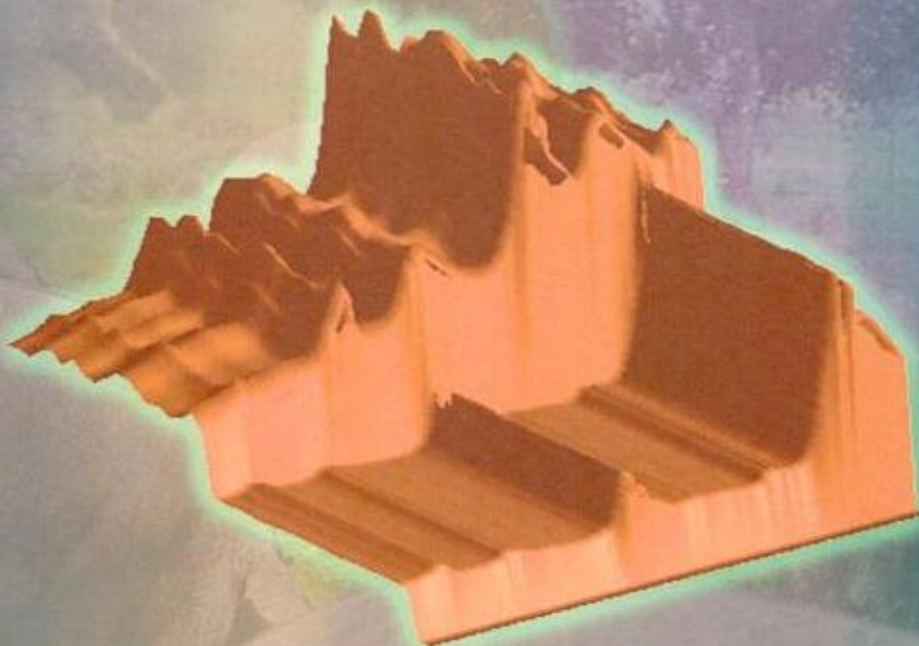




UNIVERSIDAD DE CORDOBA

**Departamento de Bromatología y
Tecnología de los Alimentos**



TESIS DOCTORAL

**Clasificación y autenticación de canales
de Cerdo Ibérico mediante Espectroscopía
en el Infrarrojo Cercano (NIRS)**

Juan García Olmo
2002

TITULO: *Clasificación y autenticación de canales de cerdo ibérico mediante espectroscopía en el infrarrojo cercano (NIRS)*

AUTOR: *JUAN GARCIA OLMO*

© Edita: Servicio de Publicaciones de la Universidad de Córdoba. 2009
Campus de Rabanales
Ctra. Nacional IV, Km. 396
14071 Córdoba

www.uco.es/publicaciones
publicaciones@uco.es

ISBN-13: 978-84-7801-993-9
D.L.: CO 1290-2009

UNIVERSIDAD DE CÓRDOBA

**DEPARTAMENTO DE BROMATOLOGÍA Y
TECNOLOGÍA DE LOS ALIMENTOS**

TESIS DOCTORAL

**CLASIFICACIÓN Y AUTENTIFICACIÓN DE CANALES
DE CERDO IBÉRICO MEDIANTE ESPECTROSCOPIA
EN EL INFRARROJO CERCANO (NIRS)**

JUAN GARCÍA OLMO

2002

UNIVERSIDAD DE CÓRDOBA

**DEPARTAMENTO DE BROMATOLOGÍA Y TECNOLOGÍA DE LOS
ALIMENTOS**

TESIS DOCTORAL

**Clasificación y autenticación de canales de cerdo Ibérico
mediante espectroscopía en el infrarrojo cercano (NIRS)**

Memoria presentada por

JUAN GARCÍA OLMO

para optar al título de Doctor en Ciencias Químicas

Vº Bº Directores

Prof. Dra. Ana Garrido Varo

Prof. Dr. Emiliano J. De Pedro Sanz

Ana Garrido Varo, Profesora Titular del Departamento de Producción Animal de la Universidad de Córdoba

INFORMA

que la presente Tesis Doctoral titulada “Clasificación y autenticación de canales de cerdo Ibérico mediante espectroscopía en el infrarrojo cercano (NIRS)”, ha sido realizada bajo su dirección por el Licenciado Juan García Olmo y reúne las condiciones exigidas para optar al grado de Doctor en Ciencias Químicas, por lo que considera procede su presentación.

Córdoba, octubre de 2002

Ana Garrido Varo

Emiliano Jesús De Pedro Sanz, Profesor Titular del Departamento de Producción Animal de la Universidad de Córdoba

INFORMA

que la presente Tesis Doctoral titulada “Clasificación y autenticación de canales de cerdo Ibérico mediante espectroscopía en el infrarrojo cercano (NIRS)”, ha sido realizada bajo su dirección por el Licenciado Juan García Olmo y reúne las condiciones exigidas para optar al grado de Doctor en Ciencias Químicas, por lo que considera procede su presentación.

Córdoba, octubre de 2002

Emiliano Jesús De Pedro Sanz

M^a Teresa Sánchez Pineda de las Infantas, Profesora Titular del Departamento de Bromatología y Tecnología de los Alimentos de la Universidad de Córdoba

INFORMA

que la presente Tesis Doctoral titulada “Clasificación y autenticación de canales de cerdo Ibérico mediante espectroscopía en el infrarrojo cercano (NIRS)”, ha sido realizada bajo su tutoría por el Licenciado Juan García Olmo y, de acuerdo con los informes de sus Directores, reúne las condiciones exigidas para optar al grado de Doctor en Ciencias Químicas, por lo que considera procede su presentación.

Córdoba, octubre de 2002

M^a Teresa Sánchez Pineda de las Infantas

***Lo único que podemos decir
con certeza del futuro es que
éste es difícil de predecir***

(S. Wold y M. Sjöström)

AGRADECIMIENTOS

Me resulta difícil expresar la enorme satisfacción personal que poseo al escribir el primer, si bien cronológicamente último, apartado de esta Tesis. Satisfacción por alcanzar aquella meta que me propuse al comenzar mi formación universitaria, como era el iniciarme en el apasionante mundo de la investigación, realizando un trabajo eminentemente práctico en el que se abordara y se intentaran proponer soluciones reales a algún problema expuesto por la propia sociedad, logrando con ello el máximo grado de formación académica.

El recorrido hasta aquí no ha sido, en absoluto, nada fácil. Ha sido demasiado largo, con un difícil comienzo, con quizá demasiada dedicación durante su desarrollo a otra multitud de proyectos además de al propio trabajo de investigación y con un final verdaderamente agónico e interminable, tanto para mí como para las personas que se encontraban próximas en el terreno personal y laboral. Sin embargo, he de reconocer que todo el tiempo empleado y el trabajo realizado para esta Tesis han merecido la pena.

Durante todo este recorrido he recibido apoyo y ayuda de numerosas personas por lo que considero fundamental el hacerles una pequeña mención al inicio de este trabajo, en señal de agradecimiento.

Si alguien ha trabajado intensamente para que pudiera llegar a este punto ellos han sido Ana y Emiliano. Además de codirectores y, desde mi punto de vista, coautores también de este trabajo, he tenido la gran suerte de poder recibir algunos de sus profundos conocimientos y muchos de sus buenos consejos. Siempre han estado dispuestos a dedicarme algo de su tiempo, incluso en las peores rachas, a pesar de mis continuas preguntas y prisas de última hora. Sin lugar a dudas, han sido mejores profesores que yo alumno.

Si todo son elogios desde un punto de vista profesional, en el aspecto personal no se puede pedir más. Han sabido crear el ambiente ideal entre el grupo, animarme en los peores momentos e incluso me han llamado la atención, casi siempre con razón, por alguna de mis cabezonerías. Creo que voy a echar de menos todos los buenos momentos que hemos tenido durante estos años y desearía que se repitieran en el futuro.

Todo este trabajo de investigación se ha realizado en el Dpto. de Producción Animal de la Escuela Técnica Superior de Ingenieros Agrónomos y Montes y, en concreto, dentro del Grupo de Investigación "Zootecnia". Quiero dejar constancia de la

enorme calidad profesional y humana de todos los integrantes de este Grupo y de su gran capacidad de trabajo.

Gracias a Augusto, José Emilio, Juan Manuel, Miguel, Pablo, entre otras cosas, por intentar responder a mis numerosas preguntas. Gracias a Mari y Mari Carmen por estar siempre dispuestas a echar una mano. Gracias Paquita por tu amabilidad, por tus detalles horneados y por analizar la muestra de grasa (no sabes lo útil que me ha sido). Gracias Antonio y Alberto por ayudar incluso antes de que se os pida y por hacer más agradable el día a día del laboratorio. Gracias Víctor, Nieves, Lola, Pastora, Leo, Mari Jose, Jara, Juanma (¿se me olvida alguien?) por crear un buen ambiente de trabajo, por ayudarme en alguna que otra duda y por aguantarme en mis días malos. Para los que estén con la Tesis, ánimo que al final todo se acaba. Gracias a todos los alumnos con los que he colaborado en el Dpto. porque junto con vosotros he avanzado un poco más en el conocimiento de la tecnología NIRS y del cerdo Ibérico.

A la Universidad de Córdoba, al Servicio Central de Apoyo a la Investigación y al Excmo. Ayuntamiento de Pozoblanco, a la D.O. “Jamón de Huelva”, al personal del Laboratorio Agroalimentario Córdoba de la Consejería de Agricultura de la Junta de Andalucía, a A.E.C.E.R.I.B.E.R., a COVAP, a Sánchez Romero Carvajal-Jabugo he de agradecerles el haberme facilitado las muestras, los medios y la financiación para poder realizar este trabajo.

A mi familia y a todos mis amigos/as, tanto de Pozoblanco como de Córdoba. Gracias por soportar alguna que otra charla demasiado extensa sobre la Tesis y por entender que este trabajo de investigación no me ha permitido dedicarles más tiempo a ellos.

Por último, y no por ello menos importante, he de agradecer todo su apoyo a los dos pilares en los que me he apoyado para mantener mi ilusión por terminar esta Tesis.

Gracias papa, gracias mama por haber hecho posible que llegara hasta aquí, por apoyarme y animarme en todo momento (incluso en los más difíciles para todos) y por tener la suficiente paciencia y la total confianza conmigo cuando os decía que esto tendría su fin.

Olimpia, tu han sido quien más ha tenido que soportar la Tesis y sabes la satisfacción que supone para mi el haber llegado a este punto. Gracias por tu ánimo, por tu apoyo, por tu ayuda, por esperarme, por soportarme, por tu paciencia, ... por todo. Espero poder recompensarte con creces por el tiempo perdido.

A Mis Padres.

A Olimpia.

ÍNDICE

1	INTRODUCCIÓN Y OBJETIVOS	1
2	REVISIÓN BIBLIOGRÁFICA	11
2.1	MÉTODOS DE EVALUACIÓN DEL RÉGIMEN ALIMENTICIO EN CANALES DE CERDO IBÉRICO	13
2.1.1	<i>Sensación al tacto</i>	14
2.1.2	<i>Puntos de fusión y deslizamiento</i>	15
2.1.3	<i>Ácidos grasos</i>	17
2.1.4	<i>Otros métodos de evaluación</i>	22
2.1.4.1	Triglicéridos en grasa	22
2.1.4.2	Análisis de la fracción insaponificable de la grasa	23
2.1.4.3	Antioxidantes en músculo	24
2.1.4.4	Análisis de isótopos	25
2.1.4.5	Análisis mediante olfatometría electrónica	25
2.2	TECNOLOGÍA NIRS	27
2.2.1	<i>Bases teóricas de la absorción NIR</i>	28
2.2.2	<i>Instrumentación y modos de análisis</i>	31
2.2.3	<i>Quimiometría y su aplicación en la tecnología NIRS</i>	35
2.2.3.1	Análisis cuantitativo	38
2.2.3.2	Análisis cualitativo	46
2.2.3.3	Clonación / estandarización de instrumentos y transferencia de calibraciones	52
2.3	APLICACIONES DE LA TECNOLOGÍA NIRS EN GRASAS Y ACEITES	57
2.3.1	<i>Análisis espectral de asignación de bandas</i>	58
2.3.2	<i>Análisis cuantitativo</i>	62
2.3.3	<i>Análisis cualitativo</i>	72
2.3.3.1	Aceites vegetales	72
2.3.3.2	Grasas	79
3	MATERIAL Y MÉTODOS	85
3.1	OPTIMIZACIÓN DE LA METODOLOGÍA PARA LA DETERMINACIÓN DE ÁCIDOS GRASOS EN CANALES DE CERDO IBÉRICO	87
3.1.1	<i>Estudio de localización anatómica de toma de muestras</i>	87
3.1.1.1	Muestras de tejido adiposo empleadas	87
3.1.1.2	Metodología de obtención de grasa líquida	88
3.1.1.3	Determinación de ácidos grasos por cromatografía de gases	89
3.1.2	<i>Estudio del número mínimo de muestras de grasa representativas de la media de un lote</i>	90
3.1.3	<i>Ensayo colaborativo</i>	92
3.1.3.1	Preparación y análisis de muestras	92
3.1.3.2	Tratamiento estadístico de resultados	93
3.2	PUESTA A PUNTO DE METODOLOGÍAS PARA OBTENCIÓN DE INFORMACIÓN ESPECTROSCÓPICA NIR EN GRASA	101
3.2.1	<i>Modalidad de doble transmisión</i>	101
3.2.1.1	Muestras	101
3.2.1.2	Análisis NIRS	101
3.2.1.3	Cálculo de la repetibilidad espectral	102
3.2.2	<i>Modalidad de interactancia-reflectancia</i>	105
3.2.2.1	Muestras	105
3.2.2.2	Análisis NIRS	105
3.2.2.3	Cálculo de la repetibilidad espectral	106
3.2.3	<i>Influencia de la congelación de muestras de grasa de cerdo Ibérico sobre su información espectral NIR</i>	108
3.2.3.1	Muestras	108
3.2.3.2	Análisis NIRS	109
3.2.3.3	Tratamiento quimiométrico de datos espectrales	110
3.3	DESARROLLO Y EVALUACIÓN DE CALIBRACIONES NIRS PARA LA PREDICCIÓN DEL CONTENIDO EN ÁCIDOS GRASOS	113
3.3.1	<i>Modalidad de doble transmisión</i>	113
3.3.1.1	Muestras	113
3.3.1.2	Análisis químico de referencia	114
3.3.1.3	Obtención de datos espectroscópicos NIR	114

3.3.1.4	Tratamiento quimiométrico de datos NIRS.....	115
3.3.2	<i>Modalidad de interactancia-reflectancia</i>	124
3.3.2.1	Muestras	124
3.3.2.2	Análisis químico de referencia.....	124
3.3.2.3	Obtención de datos espectroscópicos NIR.....	124
3.3.2.4	Tratamiento quimiométrico de datos NIRS.....	125
3.3.3	<i>Control y diagnósticos de rutina de la instrumentación NIRS utilizada</i>	125
3.3.3.1	Repetibilidad de la señal espectroscópica NIR	126
3.3.3.2	Exactitud de la longitud de onda	127
3.3.3.3	Respuesta de los detectores del instrumento.....	128
3.3.3.4	Muestra de referencia (Check Cell).....	129
3.4	DESARROLLO Y EVALUACIÓN DE MODELOS MULTIVARIANTES DE CLASIFICACIÓN DE GRASA DE CERDO IBÉRICO	132
3.4.1	<i>Material experimental</i>	132
3.4.1.1	Muestras	132
3.4.1.2	Obtención de datos espectroscópicos NIRS.....	135
3.4.1.3	Tratamiento quimiométrico de datos espectrales previo al desarrollo de modelos multivariantes	136
3.4.2	<i>Algoritmos de clasificación</i>	137
3.4.2.1	Análisis discriminante.....	137
3.4.2.2	Análisis SIMCA.....	142
3.4.2.3	Análisis discriminante basado en regresión RMCP	145
3.5	CLONACIÓN/ ESTANDARIZACIÓN DE INSTRUMENTOS Y TRANSFERENCIA DE CALIBRACIONES..	149
3.5.1	<i>Ecuaciones de calibración NIRS transferidas</i>	150
3.5.2	<i>Muestras empleadas en la estandarización</i>	150
3.5.3	<i>Obtención de espectros para muestras de estandarización y tratamiento de datos</i>	152
4	RESULTADOS Y DISCUSIÓN.....	157
4.1	OPTIMIZACIÓN DE LA METODOLOGÍA DE ANÁLISIS DE ÁCIDOS GRASOS EN CANALES DE CERDO IBÉRICO	159
4.1.1	<i>Estudio de localización anatómica de toma de muestras</i>	159
4.1.2	<i>Estudio del número mínimo de muestras representativas</i>	161
4.1.3	<i>Ensayo colaborativo</i>	168
4.2	PUESTA A PUNTO DE METODOLOGÍAS PARA OBTENCIÓN DE INFORMACIÓN ESPECTROSCÓPICA NIR EN GRASA FUNDIDA Y TEJIDO ADIPOSO.....	179
4.2.1	<i>Modalidad de doble transmisión</i>	179
4.2.2	<i>Modalidad de interactancia-reflectancia</i>	187
4.2.3	<i>Influencia de la congelación de muestras de grasa de cerdo Ibérico sobre su información espectral NIR</i>	193
4.3	DESARROLLO Y EVALUACIÓN DE ECUACIONES DE CALIBRACIÓN NIRS PARA LA PREDICCIÓN DEL CONTENIDO EN ÁCIDOS GRASOS	200
4.3.1	<i>Modalidad de doble transmisión</i>	200
4.3.1.1	Obtención de ecuaciones preliminares para la determinación del contenido en ácidos grasos en grasa fundida.....	200
4.3.1.2	Comparación de ecuaciones obtenidas con diferentes métodos de regresión: RMCPM vs. RLM	207
4.3.1.3	Comparación de ecuaciones obtenidas mediante diferentes tratamientos de derivadas	214
4.3.1.4	Predicción de colectivos de validación de diferentes periodos	220
4.3.1.5	Control y diagnósticos de rutina de la instrumentación NIRS utilizada	224
4.3.1.6	Aplicación de ecuaciones obtenidas con diferentes tratamientos de derivadas a colectivos de validación de diferentes periodos.....	233
4.3.1.7	Evaluación del pretratamiento de derivadas sobre al cálculo del estadístico H	247
4.3.1.8	Obtención de ecuaciones para la determinación de ácidos grasos minoritarios	262
4.3.2	<i>Modalidad de interactancia-reflectancia</i>	268
4.3.2.1	Obtención de ecuaciones para la determinación del contenido en ácidos grasos en tejido adiposo	268
4.3.2.2	Comparación de resultados mediante diferentes métodos de regresión: RMCPM vs. RLM.....	277
4.3.2.3	Comparación de resultados mediante diferentes tratamientos de derivadas	280
4.3.2.4	Obtención de ecuaciones para la determinación de ácidos grasos minoritarios	285
4.4	DESARROLLO Y EVALUACIÓN DE MODELOS MULTIVARIANTES DE CLASIFICACIÓN DE GRASA DE CERDO IBÉRICO	290
4.4.1	<i>Predicción del contenido en ácidos grasos</i>	291
4.4.2	<i>Desarrollo de modelos cualitativos mediante análisis discriminante</i>	296
4.4.2.1	Análisis discriminante canónico.....	297

4.4.2.2	Análisis discriminante cuadrático	303
4.4.2.3	Análisis discriminante lineal	308
4.4.3	<i>Desarrollo de modelos cualitativos mediante análisis SIMCA</i>	<i>314</i>
4.4.4	<i>Desarrollo de modelos cualitativos mediante análisis discriminante basado en regresión RMCP 318</i>	
4.4.5	<i>Selección y validación de modelos</i>	<i>324</i>
4.5	CLONACIÓN/ ESTANDARIZACIÓN DE INSTRUMENTOS Y TRANSFERENCIA DE CALIBRACIONES...	328
4.5.1	<i>Instrumentos NIRS ubicados en un mismo laboratorio</i>	<i>328</i>
4.5.1.1	Ajuste espectral	328
4.5.1.2	Transferencia de ecuaciones.....	331
4.5.2	<i>Instrumentos NIRS ubicados en laboratorios diferentes</i>	<i>337</i>
4.5.3	<i>Repetibilidad y reproducibilidad de la predicción de ácidos grasos mediante NIRS</i>	<i>341</i>
5	CONCLUSIONES.....	347
6	BIBLIOGRAFÍA.....	353

ABREVIATURAS

- ACP: Análisis de componentes principales.
- ADC: Análisis Discriminante Cuadrático.
- ADL: Análisis Discriminante Lineal.
- AECERIBER: Asociación Española de Criadores de Ganado Porcino Tronco Ibérico.
- AOCS: American Oil Chemists' Society.
- AOTF: Infrared Acousto-Optic Tunable Filter; filtros ópticos-acústicos modulables.
- ASICI: Asociación Interprofesional del cerdo Ibérico.
- ASTM: American Society for Testing and Materials.
- BOE: Boletín Oficial del Estado.
- C14:0: Ácido mirístico.
- C16:0: Ácido palmítico.
- C16:1: Ácido palmitoleico.
- C17:0: Ácido margárico.
- C17:1: Ácido margaroleico.
- C18:0: Ácido esteárico.
- C18:1: Ácido oleico.
- C18:2: Ácido lineleico.
- C18:3: Ácido linoléico.
- C20:0: Ácido araquídico.
- C20:1: Ácido gadoleico.
- CAN: variable canónica.
- CG: Cromatografía de Gases.
- COVAP: Sociedad Cooperativa Andaluza del Valle de los Pedroches.
- CP: Componente Principal.
- CTH: Contrato Tipo Homologado de compra-venta de cerdos Ibéricos.
- D.O.: Denominación de Origen.
- DESIR: Dry extract spectroscopy by infrared reflection.
- DS: Direct Standardisation.
- DT: Desviación Típica.
- DTR: Detrending (pretratamiento de corrección del fenómeno *scatter*).
- ETC: Error Típico de Calibración.
- ETD: Error típico de las diferencias.
- ETL: Error Típico del Laboratorio (método de referencia).
- ETP(C): Error Típico de Predicción Corregido para el Sesgo.
- ETP: Error Típico de Predicción.
- ETSIAM: Escuela Técnica Superior de Ingenieros Agrónomos y de Montes.
- ETVC: Error Típico de Validación Cruzada.

FIR: Far Infrared; Infrarrojo lejano.

H: Distancia H, estadístico análogo a la Distancia de Mahalanobis.

h: Estadístico h de Mandel.

HPLC: High Performance Liquid Chromatography; Cromatografía líquida de alta resolución.

IREM: Infrared Emitting Diodes; diodos emisores de luz infrarroja.

k: Estadístico k de Mandel.

KNN: K nearest neighbours.

MAPA: Ministerio de Agricultura, Pesca y Alimentación.

MEAN: Media del estadístico RMS.

MIR: Mid Infrared; Infrarrojo medio.

MSC: Multiplicative Scatter Correction (pretratamiento de corrección del fenómeno *scatter*).

NIR: Near Infrared; Infrarrojo cercano.

NIRS: Near Infrared Spectroscopy; Espectroscopía en el infrarrojo cercano.

nm: nanómetros.

OSC: Orthogonal Signal Correction (pretratamiento de corrección del fenómeno *scatter*).

PDS: Piecewise Direct Standardisation.

r: Repetibilidad.

R: Reproducibilidad.

R: Coeficiente de Correlación.

R^2 : Coeficiente de Determinación de la etapa de Calibración.

r^2 : Coeficiente de Determinación de la etapa de Validación Cruzada.

RCP: Regresión en Componentes Principales.

RER: Cociente entre el intervalo de composición de los datos de referencia y el ETVC.

RLM: Regresión Lineal Múltiple.

RMCP: Regresión en Mínimos Cuadrados Parciales.

RMCPM: Regresión en Mínimos Cuadrados Parciales Modificada.

RMS(C): Raíz cuadrada del Cuadrado Medio del Error Corregido para el Sesgo.

RMS: Raíz cuadrada del Cuadrado Medio del Error.

RPD: Cociente entre la desviación típica de composición de los datos de referencia y el ETVC.

RSD_r : Desviación Estandar Relativa de la Repetibilidad.

RSD_R : Desviación Estandar Relativa de la Reproducibilidad.

S^2 : Varianza residual.

S_L : Desviación Típica del error entre laboratorios.

S_r : Desviación Típica de la Repetibilidad.

S_R : Desviación Típica de la Reproducibilidad.

SIMCA: Soft Independent Modelling of Class Analogies.

SNV: Standard Normal Variate (pretratamiento de corrección del fenómeno *scatter*).

STD: Desviación Típica del estadístico RMS.

T: Estadístico T, análogo a una t de Student.

UCO: Universidad de Córdoba.

ÍNDICE DE TABLAS

Tabla 2.1. Criterios de clasificación establecidos en el Contrato Tipo Homologado en diversas campañas.....	18
Tabla 2.2. Aplicaciones NIR en diferentes tipos de grasas y aceites (adap. de Garrido <i>et al.</i> , en prensa)	63
Tabla 3.1. Diseño experimental del estudio de conservación de grasa	109
Tabla 3.2. Opciones incluidas en la secuencia automática de calibración (AUTOCAL).....	122
Tabla 3.3. Estadísticos de ecuaciones para el análisis NIRS de la muestra de referencia (Check Cell).....	130
Tabla 3.4. Resumen del tipo de alimentación durante el cebo de los lotes empleados para el desarrollo de los modelos	134
Tabla 3.5. Resumen del tipo de alimentación durante el cebo de los lotes empleados para validación	134
Tabla 3.6. Media, desviación típica e intervalo de los ácidos grasos a considerar en el colectivo de calibración (N=341)	150
Tabla 3.7. Composición en ácidos grasos de los colectivos de estandarización A y B así como del colectivo de validación.....	151
Tabla 4.1.1. Composición media en ácidos grasos (%) de grasa subcutánea de cerdos ibéricos, según la zona de toma de muestras para el análisis.....	160
Tabla 4.1.2. Número de animales, valores medios, mínimos y máximos de los porcentajes de ácidos grasos mayoritarios en grasa subcutánea de cerdo Ibérico de cada lote experimental y su clasificación comercial.....	161
Tabla 4.1.3. Valores medios del contenido en ácidos grasos de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote I (n=37).....	163
Tabla 4.1.4. Valores medios del contenido en ácidos grasos de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote II (n=32).....	163
Tabla 4.1.5. Valores medios del contenido en ácidos grasos de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote III (n=80).....	164
Tabla 4.1.6. Valores medios del contenido en ácidos grasos de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote IV (n=36).....	164
Tabla 4.1.7. Valores medios del contenido en ácidos grasos de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote V (n=45).....	165
Tabla 4.1.8. Valores medios del contenido en ácidos grasos de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote VI (n=71).....	165
Tabla 4.1.9. Clasificación de cada lote, en función de su composición media y número de animales clasificados en cada categoría, al aplicar los criterios del CTH de forma individualizada.....	167
Tabla 4.1.10. Media y desviación típica de ácidos grasos en las muestras analizadas por cada laboratorio.....	169
Tabla 4.1.11. Media y desviación típica de ácidos grasos en las muestras según el método de extracción de grasa empleado	171
Tabla 4.1.12. Muestras anómalas detectadas tras el Test de Cochran.....	174
Tabla 4.1.13. Muestras anómalas detectadas tras el Test de Grubbs simple y doble.....	174
Tabla 4.1.14. Parámetros estadísticos de repetibilidad y reproducibilidad.....	177
Tabla 4.2.1. Valores medios de estadísticos MEAN y STD entre espectros de una misma cápsula con grasa en estado líquido y pastoso.....	180
Tabla 4.2.2. Valores máximos, mínimos y medios de MEAN y STD entre submuestras de la misma muestra.....	182

Tabla 4.2.3. Valores Medios de RMS entre espectros de grasa analizada en cápsulas con diferentes proporción de burbujas de aire en su interior.....	185
Tabla 4.2.4. Valores de RMS en una misma posición de la sonda en tejido adiposo.....	187
Tabla 4.2.5. Valores de RMS entre diferentes posiciones, en capa externa, de la sonda en tejido adiposo	189
Tabla 4.2.6. Valores de RMS entre diferentes posiciones, en capa interna, de la sonda en la muestra ..	189
Tabla 4.2.7. Valores de RMS entre posiciones de la sonda en diferentes capas de tejido adiposo (Externa vs. Interna).....	189
Tabla 4.2.8. Influencia de la luz ambiental en los valores de RMS.....	192
Tabla 4.2.9. RMS(C) en el rango Visible y NIR entre poblaciones espectrales con diferente tiempo de conservación bajo congelación.....	194
Tabla 4.2.10. H_{VIS} y H_{NIR} entre poblaciones espectrales con diferente tiempo de conservación bajo congelación.....	196
Tabla 4.2.11. RMS(C) en el rango Visible y NIR entre poblaciones espectrales con diferentes tratamientos de congelación – descongelación.....	198
Tabla 4.2.12. H_{VIS} y H_{NIR} entre poblaciones espectrales con diferente tratamiento de congelación – descongelación.....	198
Tabla 4.3.1. Composición en ácidos grasos de las muestras empleadas como colectivo de calibración (n=341).....	202
Tabla 4.3.2. Estadísticos de ecuaciones obtenidas a partir del colectivo de calibración (n=341).....	204
Tabla 4.3.3. Composición en ácidos grasos del colectivo de validación A (n=20).....	206
Tabla 4.3.4. Estadísticos de validación externa del colectivo A (n=20).....	206
Tabla 4.3.5. Estadísticos de ecuaciones de calibración generadas mediante RLM, empleando 10 longitudes de onda por variable.....	208
Tabla 4.3.6. Longitudes de ondas seleccionadas para ecuaciones generadas mediante RLM.....	209
Tabla 4.3.7. Estadísticos de ecuaciones de calibración generadas mediante RLM, empleando 2 longitudes de onda por variable.....	210
Tabla 4.3.8. Longitudes de ondas seleccionadas para ecuaciones generadas mediante RLM.....	211
Tabla 4.3.9. Estadísticos RPD y RER de las ecuaciones desarrolladas a partir de diferentes tratamientos de derivadas.....	216
Tabla 4.3.10. Valores de ETVC máximos, mínimos y diferencias máximas, en este estadístico, para las ecuaciones de cada ácido graso.....	218
Tabla 4.3.11. Estadísticos de predicción del colectivo A con ecuaciones obtenidas mediante diferente tratamiento de derivadas.....	219
Tabla 4.3.12. Composición en ácidos grasos de las muestras del colectivo de validación B (n=30).....	220
Tabla 4.3.13. Composición en ácidos grasos de la muestra del colectivo de validación C (n=150).....	221
Tabla 4.3.14. Estadísticos de validación externa del colectivo B.....	222
Tabla 4.3.15. Estadísticos de validación externa del colectivo C.....	222
Tabla 4.3.16. Valores de \bar{y}_{mr} y σ_{mr} para los constituyentes de la muestra de referencia.....	229
Tabla 4.3.17. Estadísticos de ecuaciones obtenidas a partir del colectivo de calibración (n=341).....	236
Tabla 4.3.18. Estadísticos de predicción del colectivo A con mejores ecuaciones.....	242
Tabla 4.3.19. Estadísticos de predicción del colectivo B con mejores ecuaciones.....	242
Tabla 4.3.20. Estadísticos de predicción del colectivo C con mejores ecuaciones.....	242

Tabla 4.3.21. Valores del estadístico H de las muestras anómalas eliminadas tras los sucesivos análisis de componentes principales, para los diferentes tratamientos de derivadas.....	250
Tabla 4.3.22. Valores medios del estadístico H para los colectivos de validación tras emplear diferentes espacios de componentes principales.....	255
Tabla 4.3.23. Valores del estadístico H de las muestras anómalas eliminadas tras los sucesivos análisis de componentes principales, para los diferentes tratamientos de derivadas.....	260
Tabla 4.3.24. Valores medios del estadístico H para los colectivos de validación tras emplear diferentes espacios de componentes principales.....	261
Tabla 4.3.25. Composición en ácidos grasos minoritarios de las muestras empleadas como colectivo de calibración (n=341).....	263
Tabla 4.3.26. Estadísticos de ecuaciones NIRS para ácidos grasos minoritarios obtenidas a partir del colectivo de calibración (n=341).....	265
Tabla 4.3.27. Composición en ácidos grasos de las muestras empleadas como colectivo de calibración (n=164).....	268
Tabla 4.3.28. Estadísticos de ecuaciones obtenidas a partir del colectivo de calibración (n=164).....	270
Tabla 4.3.29. Composición en ácidos grasos del colectivo de validación D (n=15).....	274
Tabla 4.3.30. Estadísticos de validación externa del colectivo D (n=15).....	275
Tabla 4.3.31. Estadísticos de ecuaciones de calibración generadas mediante RLM, empleando 10 longitudes de onda por variable.....	277
Tabla 4.3.32. Longitudes de ondas seleccionadas para ecuaciones generadas mediante RLM.....	278
Tabla 4.3.33. Estadísticos RPD y RER de las ecuaciones desarrolladas a partir de diferentes tratamientos de derivadas.....	282
Tabla 4.3.34. Valores de ETVC máximos, mínimos y diferencias máximas en este estadístico para las ecuaciones de cada ácido graso.....	284
Tabla 4.3.35. Composición en ácidos grasos minoritarios de las muestras empleadas como colectivo de calibración (n=164).....	286
Tabla 4.3.36. Estadísticos de ecuaciones NIRS para ácidos grasos minoritarios obtenidas a partir del colectivo de calibración (n=164).....	288
Tabla 4.4.1. Composición en ácidos grasos de las muestras empleadas en el desarrollo de los modelos.....	293
Tabla 4.4.2. Resultados de los modelos desarrollados mediante análisis discriminante cuadrático, a partir de valores predichos de ácidos grasos y datos espectrales NIR.....	305
Tabla 4.4.3. Muestras mal clasificadas por los modelos generados mediante análisis discriminante cuadrático, a partir de valores predichos de ácidos grasos y datos espectrales NIR.....	306
Tabla 4.4.4. Resultados de modelos desarrollados mediante análisis discriminante lineal con los valores predichos de ácidos grasos y con datos espectrales NIR.....	309
Tabla 4.4.5. Muestras mal clasificadas por modelos generados mediante análisis discriminante lineal a partir de valores predichos de ácidos grasos y datos espectrales NIR.....	311
Tabla 4.4.6. Resultados de modelos desarrollados mediante análisis SIMCA a partir de valores predichos de ácidos grasos y datos espectrales NIR.....	315
Tabla 4.4.7. Muestras mal clasificadas por modelos generados mediante análisis SIMCA a partir de valores predichos de ácidos grasos y datos espectrales NIR.....	317
Tabla 4.4.8. Resultados de modelos desarrollados mediante análisis discriminante basado en regresión RMCP a partir de valores predichos de ácidos grasos y datos espectrales NIR.....	320
Tabla 4.4.9. Muestras mal clasificadas por modelos generados por análisis discriminante basado en regresión RMCP a partir de valores predichos de ácidos grasos y datos espectrales NIR.....	322

Tabla 4.4.10. Matriz de clasificación del colectivo de evaluación mediante modelos discriminantes basados en regresión RMCP.....	326
Tabla 4.5.1. Diagnósticos de los instrumentos Master y Satélite.....	329
Tabla 4.5.2. RMS(C) entre espectros de 2 submuestras de 10 muestras antes y después de estandarizar con 4 ficheros de estandarización diferentes.....	329
Tabla 4.5.3. Estadísticos de ecuaciones obtenidas a partir del colectivo de calibración (N=341).....	332
Tabla 4.5.4. Estadísticos para la predicción de C16:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite, antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	332
Tabla 4.5.5. Estadísticos para la predicción de C18:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite, antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	332
Tabla 4.5.6. Estadísticos para la predicción de C18:1 del colectivo de validación E (n=10) en los instrumentos Master y Satélite, antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	333
Tabla 4.5.7. Estadísticos para la predicción de C18:2 del colectivo de validación E (n=10) en los instrumentos Master y Satélite, antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	333
Tabla 4.5.8. Valores de laboratorio (referencia) y predichos para el colectivo de Validación E en los instrumentos Master y Satélite, antes y después de estandarizar con el colectivo de estandarización A.....	336
Tabla 4.5.9. Estadísticos para la predicción de C16:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Zaragoza), antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	338
Tabla 4.5.10. Estadísticos para la predicción de C18:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Zaragoza), antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	338
Tabla 4.5.11. Estadísticos para la predicción de C18:1 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Zaragoza), antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	338
Tabla 4.5.12. Estadísticos para la predicción de C18:2 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Zaragoza), antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	339
Tabla 4.5.13. Estadísticos para la predicción de C16:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Cabrils), antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	339
Tabla 4.5.14. Estadísticos para la predicción de C18:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Cabrils), antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	340
Tabla 4.5.15. Estadísticos para la predicción de C18:1 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Cabrils), antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	340
Tabla 4.5.16. Estadísticos para la predicción de C18:2 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Cabrils), antes y después de estandarizar con 4 colectivos de estandarización diferentes.....	340
Tabla 4.5.17. Valores de laboratorio (referencia) y predichos para el colectivo de Validación E en los diferentes instrumentos (Master y Satélites).....	342
Tabla 4.5.18. Parámetros estadísticos de repetibilidad y reproducibilidad.....	343

ÍNDICE DE FIGURAS

Figura 2.1. Modos de vibración del grupo $-\text{CH}_2$ (adaptado de Bertrand y Dufour, 2000).....	29
Figura 2.2. Esquema de modalidades de análisis NIRS de reflectancia, transmitancia y doble transmisión.....	33
Figura 2.3. Esquema de la modalidad de análisis NIRS de interactancia-reflectancia.....	33
Figura 3.1. Localización de las zonas de toma de muestras en la canal	88
Figura 3.2. Obtención de espectros de tejido subcutáneo de cerdo Ibérico mediante la modalidad de interactancia-reflectancia.....	106
Figura 3.3. Tejido adiposo con diferentes posiciones de la sonda de fibra óptica en la capa externa (E) e interna (I).....	107
Figura 3.4. Comparación de los valores máximos y mínimos de $\log(1/R)$ para los espectros del colectivo de estandarización B y el colectivo de validación.....	152
Figura 4.1.1. Estadísticos k de Mandel, agrupados por laboratorios	172
Figura 4.1.2. Estadísticos h de Mandel, agrupados por laboratorios	173
Figura 4.1.3. Representación de s_{ij} y s_{Rj} frente a los valores medios de los ácidos grasos	176
Figura 4.2.1. 10 espectros de una cápsula conteniendo muestra en estado líquido y pastoso	181
Figura 4.2.2. 10 espectros de 1 cápsula conteniendo muestra en estado líquido y pastoso (detalle entre 400 y 1100 nm).....	181
Figura 4.2.3. Espectros de 10 submuestras de la misma muestra.....	183
Figura 4.2.4. Espectros de 10 submuestras de la misma muestra.....	183
Figura 4.2.5. Cápsulas conteniendo grasa de cerdo con diferente contenido de burbujas de aire.....	185
Figura 4.2.6. Espectros de cápsulas con presencia de burbujas de aire.....	186
Figura 4.2.7. Detalle de espectros de cápsulas con presencia de burbujas de aire.....	186
Figura 4.2.8. 4 espectros en la posición Loc I3.....	188
Figura 4.2.9. 4 espectros en 4 posiciones diferentes de la muestra	191
Figura 4.2.10. Espectros medios en la misma posición con y sin protección de la luz ambiental	192
Figura 4.3.1. Diagramas de distribución de frecuencias del contenido en ácidos grasos en el colectivo de calibración global	202
Figura 4.3.2. Resultados de Validación al emplear ecuaciones NIRS generadas mediante RMCPM	207
Figura 4.3.3. Coeficientes de longitudes de onda seleccionadas mediante RLM en ecuaciones generadas para un máximo de 10 longitudes de onda por variable.....	209
Figura 4.3.4. Coeficientes de longitudes de onda seleccionadas mediante RLM en las ecuaciones generadas para un máximo de 2 longitudes de onda por variable.....	211
Figura 4.3.5. Resultados de validación de ecuaciones NIRS generadas mediante RLM, empleando 10 longitudes de onda por variable	213
Figura 4.3.6. Resultados de validación de ecuaciones NIRS generadas mediante RLM, empleando 2 longitudes de onda por variable	213
Figura 4.3.7. Valores de ETVC obtenidos para ecuaciones desarrolladas con diferente tratamiento de derivadas.....	215
Figura 4.3.8. Valores de r^2 obtenidos para ecuaciones desarrolladas con diferente tratamiento de derivadas.....	215
Figura 4.3.9. Diagnostico de ruido. Valores medios, máximo y mínimos de RMS(C), para el rango NIR, en el periodo de tiempo estudiado.....	226

Figura 4.3.10. Exactitud de longitud de onda para el rango NIR en el periodo de tiempo estudiado	228
Figura 4.3.11. Respuesta de la lámpara para el rango NIR en el periodo de tiempo estudiado	228
Figura 4.3.12. Predicciones de Proteína Bruta (% en MS) en muestra de referencia (Check Cell) durante el periodo considerado	230
Figura 4.3.13. Predicciones de Grasa Bruta (% en MS) en muestra de referencia (Check Cell) durante el periodo considerado	230
Figura 4.3.14. Predicciones de Fibra Bruta (% en MS) en muestra de referencia (Check Cell) durante el periodo considerado	231
Figura 4.3.15. Predicciones de Materia Seca (%) en muestra de referencia (Check Cell) durante el periodo considerado	231
Figura 4.3.16. Valores de ETP de diferentes colectivos con ecuaciones desarrolladas para C16:0 con diferente tratamiento de derivadas	234
Figura 4.3.17. Valores de ETP de diferentes colectivos con ecuaciones desarrolladas para C18:0 con diferente tratamiento de derivadas	234
Figura 4.3.18. Valores de ETP de diferentes colectivos con ecuaciones desarrolladas para C18:1 con diferente tratamiento de derivadas	235
Figura 4.3.19. Valores de ETP de diferentes colectivos con ecuaciones desarrolladas para C18:2 con diferente tratamiento de derivadas	235
Figura 4.3.20. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación seleccionada para determinar C16:0	240
Figura 4.3.21. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación seleccionada para determinar C18:0	240
Figura 4.3.22. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación seleccionada para determinar C18:1	241
Figura 4.3.23. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación seleccionada para determinar C18:2	241
Figura 4.3.24. Gráfico de control de valores predichos para C16:0 en colectivo C	245
Figura 4.3.25. Gráfico de control de valores predichos para C18:0 en colectivo C	245
Figura 4.3.26. Gráfico de control de valores predichos para C18:1 en colectivo C	246
Figura 4.3.27. Gráfico de control de valores predichos para C18:2 en colectivo C	246
Figura 4.3.28. Valores medios de H para el colectivo C calculados a partir de diferentes espacios de componentes principales	257
Figura 4.3.29. Valores medios de H para el colectivo C calculados con diferentes espacios de componentes principales	258
Figura 4.3.30. Evolución de la precisión de la longitud de onda en el rango NIR durante el periodo de análisis del colectivo C	258
Figura 4.3.31. Diagramas de distribución de frecuencias del contenido en ácidos grasos minoritarios del colectivo de calibración	264
Figura 4.3.32. Diagramas de distribución de frecuencias del contenido en ácidos grasos en el colectivo de calibración global	269
Figura 4.3.33. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación de predicción para C16:0	272
Figura 4.3.34. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación de predicción para C18:0	272
Figura 4.3.35. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación de predicción para C18:1	273

Figura 4.3.36. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación de predicción para C18:2.....	273
Figura 4.3.37. Resultados de validación al emplear ecuaciones NIRS generadas mediante RMCPM...	276
Figura 4.3.38. Coeficientes de longitudes de onda seleccionadas en ecuaciones mediante RLM con un máximo de 10 longitudes de onda por variable.....	278
Figura 4.3.39. Resultados de Validación al emplear ecuaciones NIRS generadas mediante RLM, empleando 10 longitudes de onda por variable.....	279
Figura 4.3.40. Valores de ETVC obtenidos para ecuaciones desarrolladas con diferente tratamiento de derivadas.....	281
Figura 4.3.41. Valores de r^2 obtenidos para ecuaciones desarrolladas con diferente tratamiento de derivadas.....	282
Figura 4.3.42. Valores de ETP del colectivo D obtenidos con ecuaciones desarrolladas con diferente tratamiento de derivadas.....	285
Figura 4.3.43. Diagramas de distribución de frecuencias del contenido en ácidos grasos minoritarios del colectivo de calibración.....	287
Figura 4.4.1. Representación de las muestras en los modelos generados mediante análisis discriminante canónico.....	299
Figura 4.5.1. Comparación de espectros de grasa líquida de cerdo Ibérico obtenidos en el instrumento Master y Satélite antes y después de la estandarización con colectivos A y B.....	331
Figura 4.5.2. Comparación de espectros de grasa líquida de cerdo Ibérico obtenidos en el instrumento Master y Satélite antes y después de la estandarización con cuatro colectivos diferentes.....	331

1 INTRODUCCIÓN Y OBJETIVOS

Un sector de enorme importancia para la economía española es el sector del cerdo Ibérico (Buxadé y Daza, 2000). Esta raza se puede considerar como la respuesta a un proceso selectivo, modelado por imperativos ambientales, que ha dado lugar a una de las especies mejor dotadas para el aprovechamiento de los recursos naturales del ecosistema mediterráneo de dehesa y para producir, posteriormente, piezas cárnicas selectas y de alto valor comercial. Dichas piezas y particularmente los jamones, gozan de reconocido prestigio y calidad a nivel nacional y, más recientemente, a nivel internacional (Vidal-Aragón, 1996; Ventanas, 2000).

La producción tradicional del cerdo Ibérico en su etapa final de engorde o etapa de cebo, estaba basada en el aprovechamiento de los recursos naturales existentes en el ecosistema de dehesa, donde la disponibilidad de alimentos era variable a lo largo del año, siendo el período de mayor abundancia el periodo comprendido entre los meses de noviembre y marzo. En dicho periodo, los animales se alimentan en la dehesa fundamentalmente a base de bellotas (de encina y/o alcornoque), junto con bulbos, hierba, tubérculos e incluso pequeños animales, dando lugar a un sistema de producción en extensivo característico denominado *montanera*.

Hoy en día, la amplia disponibilidad de materias primas y alimentos compuestos ha traído consigo la introducción de otros sistemas de producción de cerdo Ibérico, además del tradicionalmente empleado de *montanera*. En dichos sistemas de producción, se han introducido entre otros aspectos, nuevas fuentes de alimentación a base de cereales junto con el cruce con razas más magras y de mayor crecimiento diario que los animales genéticamente puros (Buxadé y Daza, 2000).

Los resultados de diferentes trabajos de investigación han mostrado, que si bien son muchos los factores que pueden intervenir en la calidad final de los productos del cerdo Ibérico (De Pedro, 1987; De Pedro, 2000), la alimentación recibida por el animal en el último período de cebo determina considerablemente las características finales de la canal y de los productos derivados, tanto desde el punto de vista cuantitativo como cualitativo.

De forma particular, esta diferente alimentación durante la etapa de cebo condiciona la composición y características de la grasa subcutánea, tales como la composición en ácidos grasos y triglicéridos (De Pedro, 1987; Izquierdo y Nieto, 1989;

Fallola *et al.*, 1989; De Pedro y Secondi, 1991; Flores, 1992; León, 1992; Ordóñez y de la Hoz, 1992; García-Regueiro, 1992; Vidal-Aragón, 1996) y asimismo la calidad sensorial de los productos (Vidal-Aragón, 1996; Ventanas, 2000). Otro factor de importancia que afecta a la calidad de la canal es el tipo genético, Ibérico puro ó cruzado con Duroc-Jersey (Dobao *et al.*, 1987).

Un gran reto al que se enfrentan actualmente las industrias del cerdo Ibérico y las estructuras asociativas, tales como las diferentes Denominaciones de Origen, Asociaciones o Consorcios de Promoción y/o Exportación, es el de disponer de un sistema integral de control de calidad de animales y piezas, de forma que permita garantizar en todo momento la autenticidad de los productos y permita fijar su precio en función de las diferentes calidades existentes.

Así, las diferentes Denominaciones de Origen (“Guijuelo”, “Dehesa de Extremadura”, “Jamón de Huelva” y “Los Pedroches”), industrias del sector y asociaciones de ganaderos e industriales, han implantado programas de control de la calidad, principalmente orientados a la determinación del régimen alimenticio seguido por los animales, en especial durante el periodo de montanera. Dichos programas contemplan diversos controles de campo como peso de entrada, tiempo de permanencia en montanera, cantidad de bellota, número de animales, etc., junto con resultados del análisis de composición en ácidos grasos de la grasa subcutánea.

Paralelamente, y a petición de la Asociación Interprofesional del Cerdo Ibérico (ASICI), el Ministerio de Agricultura, Pesca y Alimentación (MAPA), estableció en su día el denominado Contrato Tipo Homologado (CTH) de compraventa de cerdos Ibéricos cebados con destino a su sacrificio y elaboración. En dicho contrato, que se publica en el Boletín Oficial del Estado para cada campaña de sacrificio, se establecen las especificaciones de calidad que han de cumplir los cerdos, tanto en peso como en alimentación, así como el contenido en ácidos grasos mayoritarios (palmítico ó C16:0, esteárico ó C18:0, oleico ó C18:1 y linoleico ó C18:2) determinados por cromatografía de gases, de la grasa subcutánea de la canal, para la clasificación de animales dentro de una determinada categoría comercial.

Así, de acuerdo con la nueva norma de calidad, las categorías comerciales existentes en función del régimen alimenticio empleado durante la etapa de cebo son:

- ✓ De bellota o terminado en montanera: aquel que se destina al sacrificio inmediatamente después del aprovechamiento de la montanera, y que al menos reúna las siguientes características: 10 meses de edad mínima a la entrada en montanera, peso medio de entrada en montanera de la partida comprendido entre 80,5 y 115 kilos (7 y 10 arrobas) y con una reposición mínima en montanera de 46 kilos (4 arrobas).
- ✓ De recebo o terminado en recebo: aquel que después de reponer como mínimo 28,75 kilos (2,5 arrobas) en régimen de montanera, a la que debe haber entrado con un mínimo de 10 meses de edad, su cebo es completado hasta el momento de sacrificio mediante el aporte de piensos, constituidos fundamentalmente de cereales y leguminosas.
- ✓ De cebo (también denominado pienso): aquel animal, con una edad mínima de sacrificio de 10 meses, cuya alimentación hasta alcanzar el peso de sacrificio se basa en piensos constituidos fundamentalmente de cereales y leguminosas.

Pese a los esfuerzos realizados para la puesta en marcha de los mencionados programas de control de calidad, una limitante de orden práctico, para su implantación generalizada en el control de todos los animales producidos, se deriva del alto coste asociado, tanto a la realización de controles de campo, como de la determinación de ácidos grasos por cromatografía gaseosa.

Ello provoca que, aunque con ciertas diferencias entre las diferentes Denominaciones de Origen e industrias, los controles de campo sean escasos y los análisis se realicen sobre una única muestra representativa del muestreo realizado sobre cada partida, muestra que suele estar formada por un 20% a 50% de grasa de cerdos de cada partida, a pesar de que es probable que exista una importante variabilidad individual en cuanto a la composición de la grasa de un lote de animales. Dado el alto coste de los productos derivados de cerdo Ibérico, y en especial, sus piezas nobles como el jamón, paleta o lomo, sería deseable el poder disponer de herramientas para poder realizar un control de calidad individualizado de cada uno de los animales de un mismo lote.

Otra limitante de importancia sería, la derivada de la falta de aplicación de criterios multivariantes de clasificación de animales basados fundamentalmente en la obtención de modelos de discriminación utilizando como variables, el contenido en ácidos grasos de la grasa subcutánea de la canal. Trabajos de investigación en este terreno han mostrado, que la clasificación de canales en base a ácidos grasos considerados individualmente, no resulta adecuada para distinguir diferentes tipos de alimentación, debido a la variabilidad que los sistemas de engorde en montanera origina en la composición en ácidos grasos (Izquierdo y Nieto, 1989; De Pedro y Secondi, 1991; Osorio *et al.*, 1992; Flores, 1992; Casillas, 1995).

Una aproximación más adecuada sería la de sistemas de clasificación basados en análisis multivariante (análisis discriminante), utilizando como variables el contenido de los ácidos grasos mayoritarios en grasa subcutánea de animales sometidos a regímenes alimenticios diferentes y conocidos (Izquierdo y Nieto, 1989; Fallola *et al.*, 1989; Peralta, 1996). La utilidad de esta aproximación ha sido corroborada por De Pedro (1995), en un total de más de 1.500 cerdos cuya alimentación ha sido controlada durante más de 5 campañas consecutivas, en estrecha colaboración con industrias del sector.

Una técnica con un gran potencial para poder ser aplicada en los programas de control de calidad en el Sector del Cerdo Ibérico es la Espectroscopía de Reflectancia en el Infrarrojo Cercano (NIRS). Esta técnica ha evolucionado en menos de tres décadas, pasando de ser considerada una técnica analítica de aplicaciones muy concretas (p. ej.: cereales y harinas) y de limitado conocimiento a nivel de sus bases teóricas, a lograr una excelente reputación como técnica precisa y reproducible para el análisis cualitativo y cuantitativo en la industria agroalimentaria, farmacéutica, química, y para otras aplicaciones en medicina, medio ambiente, etc. (Osborne y Fearn, 1986; Williams y Norris, 1987a y b; Burns y Ciurczak, 1992; Osborne *et al.*, 1993; Davies y Williams, 1996).

Los datos de absorbancia de 780 a 2500 nm, obtenidos de un instrumento NIRS, pueden ser empleados para realizar un análisis cuantitativo (p. ej.: predicción de constituyentes químicos) y cualitativo (p. ej.: clasificación de grupos de calidad, identificación de muestras que se desvían de un estándar de producción, etc.) de un determinado producto. En ambos tipos de análisis es preciso desarrollar un modelo de

predicción multivariante, que relacione los datos espectrales con los atributos cuantitativos y/o cualitativos (Martens y Næs, 1989; Mark, 1992; Shenk y Westerhaus, 1995b y 1996; Williams y Sobering, 1996; Downey, 1996; Martens y Martens, 2000; Næs *et al.*, 2002).

Garrido *et al.* (1996) indican que la tecnología NIRS reúne todos los requisitos necesarios para un control de calidad moderno, a nivel de la industria agroalimentaria. Entre las ventajas atribuidas a la tecnología NIRS se encuentran las de ser una técnica rápida dada su alta velocidad de respuesta, no destructiva, posee un bajo coste analítico por muestra, con una nula o escasa necesidad de preparación de muestra, presenta una gran facilidad de uso en análisis de rutina, con posibilidad de llegar a alcanzar una precisión próxima a la del método de referencia empleado, posee una reproducibilidad mejor incluso que la del método de referencia, es una técnica multiproducto y multiatributo ya que el mismo instrumento es adecuado para diferentes productos y parámetros analíticos, no es necesario el uso de reactivos químicos ni se generan residuos contaminantes, siendo capaz de predecir atributos químicos, físicos y sensoriales.

Trabajos de investigación llevados a cabo por el Departamento de Producción Animal de la Escuela Técnica Superior de Ingenieros Agrónomos y de Montes de la Universidad de Córdoba (De Pedro *et al.*, 1992; Hervás *et al.*, 1994; Casillas, 1995) y otros posteriormente realizados en la Universidad de Extremadura (González, 1997) han demostrado la viabilidad de la técnica NIRS para la tipificación de las canales de cerdo Ibérico (De Pedro *et al.*, 1995).

Estos trabajo iniciales han sido realizados utilizando muestras de grasa, procedentes de animales producidos bajo condiciones experimentales muy controladas, lo que ha dado lugar a la obtención de ecuaciones de calibración o modelos de clasificación que en el terreno científico NIRS se denominan calibraciones o modelos de “población cerrada o específica” (Abrams, 1985; Shenk *et al.*, 1992; Shenk y Westerhaus, 1996).

Para la aceptación e implantación de la tecnología en el sector del cerdo ibérico, es necesario desarrollar trabajos que aúnen aspectos científicos y de tecnología aplicada, particularmente en aspectos tales como es la obtención de calibraciones o

modelos globales, de forma que pudieran contener la amplia variabilidad existente en la calidad de las canales procedentes de animales producidos en condiciones de campo reales. Para ello es necesario un adecuado diseño científico (Shenk y Westerhaus, 1996) y una estrecha colaboración con productores e industrias del Sector.

Dada esta situación, el Departamento de Producción Animal de la Escuela Técnica Superior de Ingenieros Agrónomos y Montes de la Universidad de Córdoba, inició en 1996 una colaboración estrecha con diferentes empresas y organismos del sector como son, entre otros, el Consejo Regulador de la Denominación de Origen “Jamón de Huelva”, la Cooperativa Andaluza del Valle de Los Pedroches (COVAP) y la Asociación Española de Criadores de Ganado Porcino Tronco Ibérico (A.E.C.E.R.I.B.E.R.) para el aporte de muestras e información de campo procedentes de diferentes campañas de sacrificio, que permitieran abordar un trabajo científico básico y de tecnología aplicada con los objetivos que se indican a continuación.

Objetivo General: Evaluación de la tecnología NIRS para la clasificación y autenticación de canales de cerdo Ibérico, en base al régimen de alimentación empleado durante la etapa de cebo.

Objetivos Específicos:

- Evaluación de errores y optimización de la metodología de toma de muestras y análisis de grasa de cerdo Ibérico para la clasificación de canales mediante la composición en ácidos grasos mayoritarios.
- Puesta a punto de metodologías para la obtención de información espectroscópica NIR en grasa de cerdo Ibérico mediante la modalidad de doble transmisión e interactancia-reflectancia.
- Desarrollo y evaluación de ecuaciones de calibración NIRS para la predicción del contenido en ácidos grasos en grasa fundida de cerdo Ibérico, mediante la modalidad de doble transmisión.
- Desarrollo y evaluación de ecuaciones de calibración NIRS para la predicción del contenido en ácidos grasos en tejido adiposo de cerdo Ibérico, mediante la modalidad de interactancia-reflectancia, empleando para ello instrumentos equipados con sonda de fibra óptica.

-
- Desarrollo y evaluación de modelos quimiométricos multivariantes para la clasificación de canales de cerdo Ibérico en base a datos de composición en ácidos grasos e información espectral NIR, de muestras de grasa fundida y tejido adiposo subcutáneo.
 - Diseño y validación de una metodología de estandarización de instrumentos NIRS y transferencia de calibraciones para la predicción del contenido en ácidos grasos en grasa fundida de cerdo Ibérico.

2 REVISIÓN BIBLIOGRÁFICA

2.1 MÉTODOS DE EVALUACIÓN DEL RÉGIMEN ALIMENTICIO EN CANALES DE CERDO IBÉRICO

Como se ha indicado anteriormente, las industrias y asociaciones de ganaderos y/o industriales del sector del cerdo Ibérico, se enfrentan actualmente al reto de disponer de herramientas de control de calidad de canales y productos transformados, de forma tal que permita garantizar en todo momento la autenticidad de sus productos, pudiendo fijar así un precio en función de las diferentes calidades existentes. En particular, uno de los factores de mayor interés en dicha etapa de control de calidad es el relativo a la determinación del régimen de alimentación empleado durante la etapa de cebo.

Dicha evaluación del régimen alimenticio que han tenido los animales en los meses anteriores a su sacrificio es difícil de conocer, ya que un control riguroso en campo de la alimentación requiere un gran esfuerzo y presenta un elevado componente de subjetividad. Además, en numerosas ocasiones, la contratación de los animales por parte del industrial se realiza pocas fechas antes de su sacrificio, sin haber podido tener un seguimiento y control de dicho régimen alimenticio (De Pedro, 2000).

Ello ha motivado que, desde que comenzaron a emplearse otros sistemas de alimentación basados en materias primas diferentes a las que el animal puede encontrar en la dehesa, se hayan investigado diferentes herramientas que permitieran evaluar de forma más objetiva dicha alimentación (López-Bote, 1998; De Pedro, 2000).

En general, los resultados obtenidos mediante dichas herramientas de análisis han sido complementarios a la información aportada por los controles de campo

realizados sobre los lotes de cerdos, teniéndose en cuenta, en la mayoría de los casos, ambos criterios (información de campo y resultado analítico) a la hora de clasificar un lote dentro de una determinada categoría comercial.

Se han propuesto diferentes metodologías y técnicas de análisis para evaluar el régimen alimenticio empleado durante el cebo de cerdos Ibéricos (Casillas, 1995; De Pedro, 2000; Ruiz y Petró, 2000). En general, dichas metodologías pretenden establecer unas diferencias entre las categorías comerciales existentes en uno o varios parámetros analíticos, en base a la información de cerdos Ibéricos sometidos a diferentes regímenes alimenticios. A continuación se expondrán de forma general dichas metodologías analíticas, junto con sus ventajas e inconvenientes en la implantación a nivel práctico.

2.1.1 SENSACIÓN AL TACTO

Inicialmente se reconocía la calidad de los productos del cerdo Ibérico por la sensación al tacto de la grasa de cobertura de los jamones. Aquellas sensaciones de dureza en la grasa se asociaban a animales cebados a base de pienso, mientras que una sensación de grasa blanda y suave al tacto era reflejo de una alimentación en régimen de montanera.

En el cerdo Ibérico, al igual que en otros animales monogástricos, la grasa presente en la dieta (grasa exógena) se depositará en el tejido adiposo, provocando una mayor concentración de ácidos grasos insaturados en la grasa de aquellos cerdos alimentados en montanera y un mayor porcentaje de ácidos grasos saturados en los animales alimentados con piensos comerciales. Esta presencia de ácidos grasos saturados en estado sólido provocará una mayor sensación de dureza en la grasa mientras que una mayor presencia de ácidos insaturados logrará un carácter mucho más oleoso y blando en dicha grasa animal, como consecuencia de las diferencias en los puntos de fusión de dichos ácidos grasos (De Pedro, 1987 y 2000; Ruiz y Petró, 2000).

Cuando sólo se disponía de cereales (como trigo o cebada) y algún tipo de leguminosa (p.ej.: guisantes, habas) para emplear como fuentes de alimentación en el cerdo Ibérico, la sensación al tacto era un método muy fiable. Sin embargo, a medida que se fue disponiendo de otras materias primas, como maíz o soja, con mayores

contenidos en ácidos grasos insaturados que los alimentos tradicionalmente empleados, las grasas de los animales fueron reduciendo su firmeza llegando a tener una sensación táctil similar o superior a la proporcionada por los animales acabados en montanera.

Además, en dicha sensación de firmeza intervienen otros factores como la cantidad de colágeno del tejido graso, la temperatura ambiente a la que están los productos, o la experiencia previa que tenga el evaluador.

Por todo ello, la sensación al tacto es considerada como una característica muy subjetiva y el error que se puede cometer, al tratar de reconocer el régimen alimenticio que han tenido los animales, es elevado. No obstante, a falta de otra técnica no destructiva para reconocer la alimentación que ha tenido el animal, se sigue utilizando la sensación al tacto de forma orientativa para clasificar canales, de modo que si la grasa de la canal se presenta consistente y dura al tacto, indica que la cantidad de bellota consumida por el cerdo ha sido nula o escasa. Sin embargo, si la grasa presenta un carácter fluido y oleoso ello no implica que los cerdos hayan tenido una alimentación basada la montanera durante su cebo.

2.1.2 PUNTOS DE FUSIÓN Y DESLIZAMIENTO.

Las primeras metodologías de análisis propiamente dichas, introducidas como herramientas de control del régimen alimenticio empleado, fueron la medición de los puntos de fusión y de deslizamiento. Dichas técnicas tuvieron una enorme repercusión a nivel práctico, llegando incluso a tener un carácter oficial durante unos años como método de evaluación de la alimentación durante el cebo (BOE, 1988).

Ambas metodologías de análisis se basan en la determinación de la temperatura de fusión de la grasa, es decir la temperatura a la cual la grasa pasa de estado sólido a líquido. Dicha temperatura de fusión no es constante y dependerá, fundamentalmente, de los ácidos grasos presentes tanto en los triglicéridos como en la fracción insaponificable de la grasa. Cada uno de los ácidos grasos existentes en este tipo de grasa presenta un valor de temperatura de fusión diferente, elevándose su valor con el elongamiento de la cadena de carbono y disminuyendo al aumentar el grado de insaturación del ácido graso.

Así, el ácido C16:0 funde a 63°C, el ácido C18:0 funde a 69°C, el C18:1 a 13,4°C, el C18:2 a -0,5°C y el C18:3 a -11°C. Por tanto, dependiendo de la composición en ácidos grasos presente en la grasa su temperatura de fusión variará. Consecuentemente, al estar formadas las grasas por una mezcla de varios glicéridos, sólidos y líquidos, no tienen un punto de fusión definido como las sustancias puras, sino que la fusión se realiza en un intervalo de temperatura (Bockisch, 1998).

La determinación de la temperatura de fusión es complicada y tediosa de realizar. Por ello, para simplificar y facilitar esta medida, se emplea otro parámetro analítico como es la temperatura de deslizamiento. En este caso, lo que se determina es el valor en el cual una muestra de grasa asciende por un tubo capilar, al irse reblandeciendo, bajo la presión de agua caliente. Su valor será siempre inferior al de la temperatura a la cual la grasa funde totalmente.

En ensayos realizados con cerdos cebados con diferentes regímenes alimenticios se ha obtenido valores comprendidos entre 27 y 30°C para la temperatura de deslizamiento de la grasa de cerdos terminados en montanera, mientras que los valores de la grasa de cerdos de recebo oscilaron entre 30 y 30,5°C y los correspondientes a grasa de cerdos terminados exclusivamente con pienso estuvieron por encima de 32°C (MAPA, 1984). Por su parte, Dobao *et al.* (1987), con cerdos Ibéricos puros, obtuvieron valores medios de temperatura de deslizamiento de 24,9°C en los animales alimentados exclusivamente con bellota, mientras que el valor medio de los animales terminados con pienso compuesto o harina de cebada fue de 26,7 y 26,5°C respectivamente. Superiores a estos valores son los obtenidos por Flores *et al.* (1988), tanto en los animales de bellota como en los de categoría de recebo o cebo.

A pesar de la sencillez de la determinación analítica de la temperatura de deslizamiento y de su mayor objetividad que la sensación al tacto, este parámetro presenta numerosas fuentes de error probablemente debidas a la zona anatómica de donde se tome la muestra, a las condiciones de almacenamiento de la grasa, a las condiciones de preparación de la muestra o a la persona que realiza la medida (Casillas, 1995; De Pedro, 2000).

Dado el diferente grado de insaturación de la grasa de los animales cebados en montanera y pienso, otro método ensayado para diferenciar el régimen de

alimentación fue el índice de yodo. Mediante esta técnica, muy aplicada en el análisis de otro tipo de grasas animales y vegetales, se determina directamente la cantidad de yodo que reacciona con 100 gramos de grasa (AOCS, 1998). Dado que dicha reacción se produce con aquellos dobles enlaces presentes en cualquier compuesto, este análisis está directamente relacionado con el grado de insaturación presente en la grasa.

Sin embargo, tanto la temperatura de deslizamiento como el índice de yodo son fácilmente modificables con la incorporación en el pienso de materias primas con alto contenido en ácidos grasos poliinsaturados (bajo punto de fusión) y obtener así temperaturas de deslizamiento e índices de yodo similares que los obtenidos alimentando los animales en montanera (De Pedro, 2000; Ruiz y Petró, 2000).

Se tendrá así una situación similar a la de la valoración por el tacto. Si un producto presenta una temperatura de deslizamiento o fusión elevada, ello indica que el régimen alimenticio empleado se ha realizado en montanera pero, si por el contrario, el valor de la temperatura es bajo no se puede indicar con seguridad el tipo de alimentación empleado.

2.1.3 ÁCIDOS GRASOS

En el cerdo Ibérico, al igual que en otras razas porcinas, la composición en ácidos grasos del tejido adiposo depende principalmente del tipo y concentración de ácidos grasos presentes en la dieta (Enser, 1995).

La influencia tan marcada del régimen alimenticio en la composición de las grasas ha dado lugar a que se haya propuesto la composición en ácidos grasos, de la grasa subcutánea de la canal, para diferenciar cerdos Ibéricos sometidos a diferentes regímenes alimenticios durante la etapa de cebo (Flores *et al.*, 1988; Izquierdo y Nieto, 1989; De Pedro y Secondi, 1991; Casillas, 1995; Peralta, 1996; De Pedro, 2000). Dicha técnica analítica es empleada en la actualidad por el sector para la evaluación del régimen alimenticio del cerdo Ibérico durante la fase de cebo.

El contenido en ácidos grasos fue introducido hacia mediados de los años 90 en el contrato tipo homologado por el Ministerio de Agricultura, Pesca y Alimentación, como criterio a emplear, junto con la información de campo (edad mínima y peso de

entrada en montanera, duración de la montanera, reposición), para la evaluación de la alimentación en cerdos ibéricos destinados a sacrificio.

En dicho contrato tipo homologado se indican los valores máximos y mínimos que han de presentar los ácidos grasos mayoritarios (C16:0, C18:0, C18:1 y C18:2) en una muestra media de grasa subcutánea, tomada en cada lote de animales objeto de dicho contrato, para que sea considerada como perteneciente a una categoría comercial de bellota, recebo o cebo. Los valores máximos y mínimos considerados para los ácidos grasos mayoritarios en los diferentes contratos tipo homologados por el MAPA se encuentran recogidos en la Tabla 2.1.

TABLA 2.1. Criterios de clasificación establecidos en el Contrato Tipo Homologado de diversas campañas de sacrificio.

CATEGORÍA	CAMPAÑA	C18:1 (%)	C16:0 (%)	C18:0 (%)	C18:2 (%)
BELLOTA	94/95	> 52	< 22	< 9	7-9
RECEBO		> 50	22-24	9-10,5	9-10
BELLOTA	95/96	> 50	< 22,5	< 9,5	6,5-10
RECEBO		> 48	< 25	< 11	< 11
BELLOTA	96/97	> 52	< 22	< 9,5	6,5-9,5
RECEBO		> 50	< 24	< 10,5	< 11
BELLOTA	97/98	> 53	< 21,5	< 9,5	< 9,5
RECEBO		> 52	< 23	< 10,5	< 10,5
BELLOTA	98/99	> 54	< 21	< 9,5	< 9,5 (*)
RECEBO		> 52	< 23	< 10,5	< 10,5
BELLOTA	99/00	> 54	< 21	< 9,5	< 9,5 (*)
RECEBO		> 52	< 23	< 10,5	< 10,5
BELLOTA	00/01	> 54	< 21	< 9,5	< 9,5 (*)
RECEBO		> 52	< 23	< 10,5	< 10,5

(*) Se aceptarán valores superiores de C18:2 hasta 10,5 siempre que la composición de C18:1 aumente en el mismo porcentaje sobre 54%

El establecimiento de unos niveles para los 4 ácidos grasos mayoritarios en los inicios del contrato tipo homologado supuso una gran aportación al sector, dada su mayor objetividad frente a las técnicas con carácter oficial anteriormente empleadas (temperatura de deslizamiento e índice de yodo). Sin embargo, hoy en día este criterio

analítico de evaluación de la alimentación está sometido a continuas críticas tanto por parte de los industriales como de los ganaderos del sector del cerdo Ibérico.

A estas críticas también podrían unirse diferentes aspectos de su aplicación que presentan ciertos inconvenientes (De Pedro, 2000) tales como:

- ✓ Los niveles de ácidos grasos para cada categoría comercial se han modificado en cada campaña, provocando que las diferentes categorías de calidad no permanezcan constantes a lo largo de diferentes campañas.
- ✓ Las categorías de calidad están definidas en función de los niveles de composición de cuatro ácidos grasos, si bien no se indica nada respecto a qué ocurre cuando únicamente alguno de ellos sobrepasa o no llega a los niveles establecidos.
- ✓ La determinación de ácidos grasos se realiza, mediante la técnica de cromatografía de gases, sobre una muestra de grasa líquida media. Esta se obtiene por extracción, de un determinado número de muestras de grasa subcutánea, tomadas al azar de las canales de la partida en la línea de matanza. Al reducir la medida a una única muestra no se contempla la variabilidad existente en la partida.
- ✓ Con los avances en nutrición animal y formulación de piensos se podría conseguir elaborar piensos que permitiesen alcanzar, en las grasas de los cerdos que los consuman, unos niveles de esos cuatro ácidos grasos similares a los que se obtienen engordando los animales en montanera.

Además, existen numerosas limitaciones a la hora de implantar la determinación de ácidos grasos en grasa subcutánea mediante cromatografía de gases en industrias o entidades del sector, las cuales deberían ser resueltas para lograr una sistemática de clasificación objetiva, fiable a utilizar por cualquier industrial, ganadero o técnico, tal como ocurre con los sistemas de clasificación de canales de porcino de capa blanca (De Pedro, 2000). Entre estas limitaciones cabe destacar el elevado tiempo y coste de análisis por muestra, la necesidad de uso del instrumental analítico por parte de personal especializado, los elevados costes de mantenimiento y la imposibilidad de realizar un análisis completo de todos los animales de cada lote.

Todas estas limitaciones hacen que esta metodología no sea viable para poder ser implantada en numerosas empresas del sector, sobre todo en las de pequeño tamaño. La forma tradicional de aplicación de los criterios del contrato tipo homologado es la de la obtención de una única muestra representativa de un lote de cerdos y su análisis por parte de laboratorios públicos o privados.

El problema de clasificación de aquellas partidas, que en su analítica sobrepasan o no llegan a los niveles establecidos para algún ácido graso, se podría resolver con la aplicación de técnicas de análisis multivariante. Con estas técnicas, la definición de las categorías se hace sobre la base del conjunto del perfil de ácidos grasos, característico de cada tipo de régimen alimenticio, y no a sus valores individuales.

Así, las técnicas estadísticas de análisis multivariante, (como puede ser el análisis factorial, discriminante o de componentes principales), se han convertido en herramientas muy útiles para conocer cómo influyen diversos factores, fundamentalmente la alimentación durante el cebo, en las características composicionales de ácidos grasos en la grasa de cerdo Ibérico (Flores *et al.*, 1988; De Pedro, 1995; Casillas, 1995; Peralta, 1996; González, 1997; Benito *et al.*, 1998; Ruiz *et al.*, 1998; López-Bote, 1998; De Pedro, 2000).

Un ejemplo práctico de la utilización de modelos multivariantes a partir de la composición de ácidos grasos de grasa subcutánea es el descrito por De Pedro (2000). En dicho trabajo se exponen los resultados de la implantación en una industria del sector de modelos multivariantes de clasificación de canales a partir del perfil completo de composición en ácidos grasos presentes en la grasa subcutánea de cerdos Ibéricos. La eficacia de estos modelos ha sido contrastada con la clasificación de varios miles de muestras procedentes de cerdos con diferentes regímenes alimenticios y sacrificados a lo largo de más de 6 campañas, obteniéndose un escaso número de diferencias entre las evaluaciones del régimen alimenticio realizadas mediante los controles de campo y mediante el uso de dichos modelos multivariantes.

En la última década, se ha profundizado en el análisis de ácidos grasos en tejidos grasos y musculares diferentes a la grasa subcutánea e incluso en fracciones de la grasa distintas a los triglicéridos (grasa saponificable).

Así, De la Hoz *et al.* (1993) y Ruiz *et al.* (1998) eligieron la grasa hepática para la evaluación del régimen alimenticio a partir de su composición en ácidos grasos. La elección de esta grasa se basa en que se trata de un tejido metabólicamente muy activo y con un contenido en fosfolípidos muy importante. Todo ello redundaba en una velocidad de recambio de los ácidos grasos mucho mayor que en el tejido subcutáneo, por lo que se supuso que podrían utilizarse para la evaluación del régimen empleado por el animal en un periodo cercano al sacrificio (Ruiz y Petró, 2000).

Los resultados obtenidos en ambos trabajos muestran que las diferencias observadas entre 3 regímenes alimenticios, correspondientes a las categorías de bellota, recebo y cebo, son del mismo orden e incluso inferiores a las observadas por otros autores en la composición de ácidos grasos de tejido subcutáneo.

Dado que los ácidos grasos presentes en los diferentes músculos son, en gran medida, responsables de las características sensoriales de los productos curados del cerdo Ibérico, Cava *et al.* (1997), Ruiz *et al.* (1998), Tejeda (1999) y Tejeda *et al.* (2002) estudiaron las variaciones de composición en ácidos grasos en diferentes piezas.

En concreto Cava *et al.* (1997) y Ruiz *et al.* (1998) analizaron en 30 muestras de músculo masetero, de un mismo número de animales sometidos a 3 regímenes alimenticios (bellota, recebo y pienso), la composición de ácidos grasos de la grasa intramuscular (tanto de la fracción de triglicéridos como de la fracción de fosfolípidos) y de la grasa total. De las diferentes fracciones analizadas, los mejores resultados de discriminación se obtuvieron en el caso del análisis de la fracción de triglicéridos de la grasa intramuscular y de la grasa total del músculo.

Sin embargo, dado que los ácidos grasos de los fosfolípidos son los que en mayor medida sufren los procesos oxidativos durante el proceso de maduración y contribuyen al perfil de compuestos volátiles, los autores subrayan la importancia del análisis de ácidos grasos de la fracción de fosfolípidos de la grasa intramuscular, si bien su poder de discriminación es menor comparado con los resultados de la fracción de triglicéridos (Cava *et al.*, 1997; Ruiz y Petró, 2000).

Tejeda (1999) y Tejeda *et al.* (2002) analizaron un mismo número de muestras de grasa intramuscular si bien, en este caso, el músculo elegido fue el biceps femoral y únicamente se consideraron dos regímenes alimenticios (bellota y pienso).

Sobre ellas analizaron el contenido en ácidos grasos en la grasa total, en la fracción de triglicéridos y en la fracción de fosfolípidos. Los resultados de discriminación obtenidos fueron similares a los descritos por Cava *et al.* (1997) y Ruiz *et al.* (1998) para el músculo masetero.

De acuerdo con Ruiz y Petró (2000), la principal ventaja que ofrece el análisis de ácidos grasos de la fracción fosfolípida de la grasa intramuscular es la de presentar dicha composición una relación directa con la calidad del producto elaborado y con su aptitud tecnológica. Sin embargo, el análisis de ácidos grasos de fosfolípidos musculares requiere un protocolo de análisis lento y complejo que lo hace poco conveniente para su aplicación como sistema rutinario de clasificación de canales en función del régimen alimenticio.

2.1.4 OTROS MÉTODOS DE EVALUACIÓN

En la última década se han empleado numerosas técnicas analíticas para evaluar la alimentación en el cerdo Ibérico tanto a partir de grasa como de músculos procedente de su canal. Entre ellas destacan la determinación del contenido en triglicéridos y de compuestos presentes en la fracción insaponificable de la grasa, de antioxidantes naturales en el músculo y el análisis mediante isótopos estables y mediante olfatometría electrónica.

2.1.4.1 *Triglicéridos en grasa*

Casillas (1995) y Díaz *et al.* (1996) evaluaron el efecto del régimen alimenticio sobre el contenido en triglicéridos de la grasa subcutánea de cerdo Ibérico. Para ello, emplearon un total de 36 muestras representativas de 4 regímenes alimenticios (desde la categoría comercial de bellota a cebo), sobre las que se determinó el contenido en 15 triglicéridos mediante cromatografía líquida de alta resolución (HPLC).

El análisis de la varianza de los resultados mostró que para la mayoría de los triglicéridos se detectaron diferencias significativas entre los 4 lotes experimentales empleados, obteniéndose unos excelentes resultados de clasificación al realizar un

análisis discriminantes a partir de todos los valores de composición de triglicéridos de los diferentes lotes.

Posteriormente, Tejeda (1999) y Tejeda *et al.* (2002) confirmaron dicha influencia de la alimentación en el contenido en triglicéridos, en este caso, de la grasa intramuscular del músculo biceps femoral.

En realidad, al analizar la composición en triglicéridos de la grasa se está analizando indirectamente la composición en ácidos grasos, los cuales se encuentran mayoritariamente formando ésteres con la glicerina. Por ello, la misma influencia de la alimentación en la composición en ácidos grasos ha de ser observada en este caso con los diferentes triglicéridos.

A pesar de que en un principio se podía pensar en una simplificación de la determinación analítica frente al análisis de ácidos grasos, dado que no es necesario la saponificación y posterior formación de los ésteres metílicos, ello no es rigurosamente cierto. De hecho, en la práctica, tanto la determinación de ácidos grasos como la de triglicéridos son similares en cuanto a tiempo de análisis y complejidad de la técnica de análisis (Ruiz y Petró, 2000). Por ello, no ha sido empleada como método de evaluación del régimen alimenticio en industrias del sector del cerdo Ibérico.

2.1.4.2 *Análisis de la fracción insaponificable de la grasa*

La fracción insaponificable de la grasa es una fracción lipídica minoritaria, representando un porcentaje inferior al 1% de la grasa total. En dicha fracción se encuentran presentes una gran variedad de compuestos como son hidrocarburos, esteroides, carotenoides o tocoferoles (Ruiz y Petró, 2000). Tejeda (1999) y Tejeda *et al.* (2001a,b) realizaron un estudio pormenorizado de dicha fracción en la grasa intramuscular del músculo biceps femoral de cerdo Ibérico, empleando para ello un reducido número de animales distribuidos en lotes con diferente régimen alimenticio.

Los resultados obtenidos mostraron, en general, la existencia de diferencias significativas en el contenido de diferentes compuestos de la fracción insaponificable entre lotes de animales alimentados en montanera y otros cebados a base de pienso comercial. Estas diferencias no fueron tan nítidas al estudiar el contenido de hidrocarburos lineales saturados de la grasa intramuscular del biceps femoral (Tejeda *et*

al., 2001a), donde únicamente se apreciaron diferencias significativas entre lotes con alimentación en montanera y a base de pienso para el hidrocarburo n-C₁₃.

Sin embargo, las diferencias entre los lotes con diferente alimentación fue más clara al estudiar la fracción de hidrocarburos ramificados presentes en dicha grasa intramuscular (Tejeda *et al.*, 2001b). En concreto, dichos autores detectaron la presencia de un hidrocarburo ramificado, como es el neofitadieno, en lotes alimentados en montanera y su total ausencia en lotes clasificados como de la categoría de cebo (Tejeda *et al.*, 2001c). Además de dichas diferencias significativas, el contenido en el mencionado compuesto presenta una correlación muy clara con la duración del periodo de alimentación en montanera. Esto se debe, según los autores a que dicho compuesto se encuentra presente en la fracción insaponificable de la hierba, la cual es ingerida por el animal durante el cebo en montanera.

Si bien estos resultados pueden ser prometedores, la necesidad de corroborar la presencia de dicho compuesto en un mayor número de muestras, la complejidad y el elevado tiempo necesario para realizar el análisis hacen muy difícil su utilización en industrias del sector (Ruiz y Petró, 2000).

2.1.4.3 Antioxidantes en músculo

La alimentación recibida por el cerdo Ibérico influye, no sólo en la composición de los ácidos grasos y su distribución en las diferentes fracciones lipídicas, sino también en la presencia de sustancias con capacidad antioxidante en las membranas musculares, fundamentalmente tocoferoles (Rey *et al.*, 1997; Cava *et al.*, 1999 y 2000). Según Rey *et al.* (1998), la presencia de un elevado contenido de tocoferoles (en concreto los alfa y gamma tocoferoles y sus derivados) en la hierba y en la bellota, serían los responsables de la presencia de una mayor contenido de estos compuestos en los lípidos de membrana de cerdos alimentados en montanera.

Estos compuestos antioxidantes están relacionados de forma estrecha con la maduración del jamón ya que son capaces de modular las reacciones de oxidación de los lípidos de la grasa intramuscular, responsable en buena parte del aroma del producto curado. Por ello, su presencia y análisis no es únicamente interesante desde un punto de vista de evaluación de la alimentación en la canal sino también como método predictivo

del comportamiento oxidativo de la materia prima durante la maduración del jamón (Ruiz y Petró, 2000).

Dado que esta metodología de análisis consiste, básicamente en una determinación de un compuesto presente en la fracción insaponificable de la grasa intramuscular mediante HPLC, sus inconvenientes son los mismos a los ya descritos en el apartado anterior.

2.1.4.4 *Análisis de isótopos.*

El uso del análisis de isótopos para caracterizar canales de cerdos Ibéricos en función de la alimentación empleada ha sido propuesto por González-Martín *et al.* (1997 y 1999). Para ello, cuantificaron la relación entre los isótopos $^{13}\text{C}/^{12}\text{C}$ en muestras de grasa subcutánea procedente de animales pertenecientes a las categorías comerciales de bellota, recebo y cebo.

Los resultados obtenidos por dichos autores mostraron que existe una relación lineal ($r = 0,982$) entre la ganancia de peso en montanera y la proporción del isótopo ^{13}C encontrado en la grasa subcutánea. Ello puede ser debido a que la relación entre los isótopos $^{13}\text{C}/^{12}\text{C}$ en la grasa está directamente relacionada con la proporción de dichos isótopos en los alimentos que ingiere.

Las posibilidades de uso de esta técnica de análisis han sido confirmadas por González-Martín *et al.* (2001) al emplear no sólo la proporción del isótopo ^{13}C del carbono sino también la del isótopo del azufre (^{34}S).

Sin embargo, los principales inconvenientes de esta técnica de análisis son el elevado coste del instrumental necesario y la complejidad del análisis que, junto con la necesidad de la confirmación de los resultados en un mayor número de muestras, hace inviable, en la actualidad, su aplicación práctica en industrias del sector.

2.1.4.5 *Análisis mediante olfatometría electrónica*

A nivel de investigación, recientemente se está estudiando el uso de la olfatometría electrónica, también denominada nariz electrónica o artificial, para la evaluación de la alimentación de cerdos Ibéricos en la etapa de cebo.

Este instrumento genera una respuesta electrónica específica, mediante una serie de sensores, ante la presencia de determinados compuestos volátiles. Una vez obtenida dicha respuesta en una serie de grupos o lotes de aprendizaje y tras la generación de modelos predictivos a partir del tratamiento quimiométrico de dicha respuesta, es posible detectar y cuantificar la presencia de compuestos volátiles e incluso obtener una estimación del aroma global de la muestra.

González-Martín *et al.* (2000) propusieron dicha técnica de análisis para la diferenciación del régimen alimenticio a partir de muestras de grasa subcutánea. Para ello emplearon 26 muestras de grasa procedente de lotes con 6 tipos de alimentación que abarcaban todas las categorías comerciales. Los resultados mostraron la posibilidad únicamente de diferenciar de forma nítida aquellos animales pertenecientes a la categoría de cebo de aquellos pertenecientes a las categorías de recebo y bellota. Los errores de clasificación, obtenidos mediante validación cruzada, fueron del 0% al emplear análisis discriminante lineal mientras que este aumentó al 28% al aplicar un análisis SIMCA.

Las posibilidades de esta técnica fueron confirmadas por Carrapiso *et al.* (2001), donde se obtuvieron mediante esta metodología de análisis errores de clasificación del 14,5% mediante análisis discriminante y del 2,2% mediante análisis por redes neuronales.

La principal ventaja que presenta esta técnica de análisis es el bajo coste, su fácil uso y la ausencia de manipulación de muestra. Sin embargo, esta técnica analítica ha surgido en los últimos años y los actuales equipos de análisis se encuentra aun en fase de desarrollo, siendo necesaria la mejora de numerosas fuentes de error instrumentales y su evaluación rigurosa en colectivos de muestras más amplios.

2.2 TECNOLOGÍA NIRS

En la década de los 60, surge un nuevo concepto de análisis de productos agroalimentarios basado en la absorción de radiación en la región del infrarrojo cercano (780 a 2500 nm), conocido como análisis NIRS, el cual deriva su nombre del acrónimo en inglés de Espectroscopía de Reflectancia en el Infrarrojo Cercano (Near InfraRed Spectroscopy).. En la década de los 70, dicho concepto se materializa en la tecnología NIRS, basada en fundamentos de espectroscopía, matemática, estadística, quimiometría y ciencias de la computación (Shenk y Westerhaus, 1995a, Kaffka y Giamarti, 1995, Stark, 1996, Geladi, 1996).

En la actualidad, esta tecnología está ampliamente difundida en numerosos campos de investigación y se encuentra ampliamente incorporada para el control de calidad en diferentes sectores como el agroalimentario, farmacéutico, químico y petroquímico, textil, ambiental, biología o de medicina (Murray, 1990; Davies y Williams, 1996; Davies y Cho, 2002).

A esta difusión e implantación en numerosos sectores y campos de investigación han contribuido de forma importante la creación de un Comité Internacional sobre dicha tecnología el cual realiza periódicamente conferencias donde se exponen los últimos avances (Iwamoto y Kawano, 1990; Biston y Bartiaux-Thill., 1991; Murray y Cowe, 1992; Hildrum *et al.*, 1992; Flinn y Batten, 1995; Davies y Williams, 1996; Davies, 1998; Davies y Giangiacomo, 2000; Davies y Cho, 2002), los numerosos artículos científicos editados en revistas específicas (*Journal of Near Infrared Spectroscopy* o *NIR News*) y asimismo los diversos libros publicados sobre las bases teóricas y aplicaciones concretas de esta tecnología (Osborne y Fearn, 1986; Williams y Norris, 1987; Burns y Ciurczak, 1992 y Osborne *et al.*, 1993).

Todo ello ha proporcionado una gran transparencia en las bases científico-técnicas de la tecnología NIRS y por tanto ha contribuido de forma importante a su aceptación científica y académica y a su expansión a nivel de centros públicos de investigación, industrias e incluso organismos oficiales de inspección (Garrido *et al.*, 1993 y 1996, Kaffka y Giarmati, 1995).

2.2.1 BASES TEÓRICAS DE LA ABSORCIÓN NIR

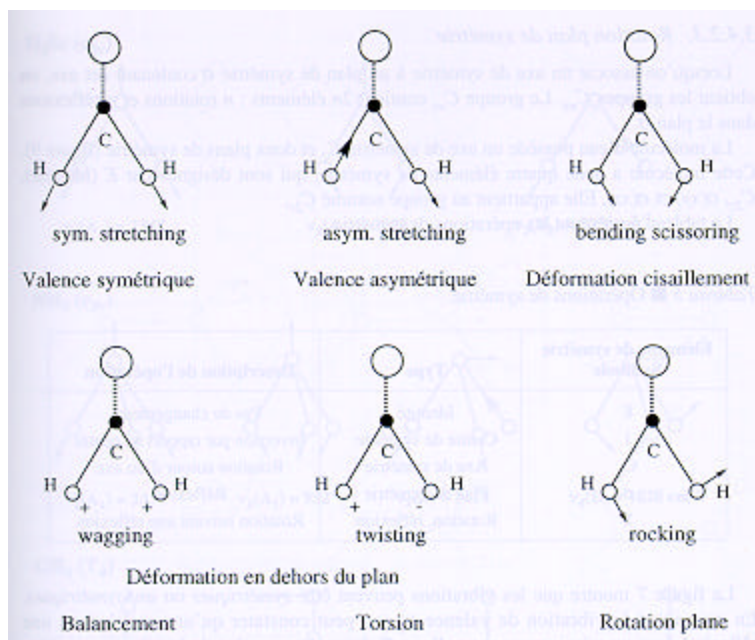
El descubrimiento de la radiación infrarroja fue realizado por Herschel (1800) cuando estudiaba el color responsable del calor de la luz solar, dentro del rango visible del espectro electromagnético. Para ello, equipado con un termómetro y un prisma con el que separar los colores de la luz solar, pudo apreciar un incremento de la temperatura al colocar el termómetro más allá de la región del color rojo del espectro visible. A esta radiación invisible la denominó radiación infrarroja (Barton, 2002).

La región del infrarrojo se puede definir como aquella sección del espectro electromagnético comprendida entre los 780 nm, final de la región del visible hasta los 25.000 nm. Dentro de dicha región es posible, a su vez, diferenciar diferentes secciones dependiendo del intervalo de longitudes de onda considerado. Así, se denomina infrarrojo medio (MIR) y lejano (FIR) a la región espectral comprendida entre 2.500 y 25.000 nm e infrarrojo cercano (NIR) a la porción de la radiación infrarroja más cercana al visible (Burns y Ciurczak, 1992). En concreto, la región NIR del espectro electromagnético se extiende desde los 780 a 2.500 nm, expresado como longitud de onda, o lo que es igual, desde 12.800 a 4.000 cm^{-1} , expresado en número de onda (Shenk y Westerhaus, 1995a; Osborne *et al.*, 1993; Næs *et al.*, 2002).

Básicamente, el fundamento de la tecnología NIRS, consiste en la emisión de un haz de luz sobre la muestra, la cual, en función de su composición, o mejor aún, de la naturaleza de los enlaces presentes en sus moléculas, fundamentalmente de aquellos del tipo -CH, -NH, -OH y -CO, interaccionará con ellos absorbiendo una determinada cantidad de radiación NIR. (Burns y Ciurczak, 1992). Los modos de vibración más importantes se suelen englobar en dos grandes grupos: aquellas vibraciones de flexión (*stretching*), tanto simétricas como asimétricas y las vibraciones

de flexión o deformación (*bending*), ya sean en el mismo plano o fuera de él. En la Figura 2.1 se representan algunos modos de vibración de un grupo $-\text{CH}_2$.

Figura 2.1. Modos de vibración del grupo $-\text{CH}_2$ (adaptado de Bertrand y Dufour, 2000).



La energía absorbida incrementa la energía de los modos de vibración de las moléculas a través de tres mecanismos diferentes (Murray y Williams, 1987; Burns y Ciurczak, 1992; Osborne *et al.*, 1993; Næs *et al.*, 2002):

- Sobretonos o armónicos de las vibraciones fundamentales existentes en la región del infrarrojo medio, los cuales son múltiplos aproximados de las vibraciones fundamentales
- Combinación de dos o más vibraciones fundamentales del infrarrojo medio
- Absorciones electrónicas, debidas al movimiento de electrones entre diferentes niveles energéticos.

La forma más usual de cuantificar la absorción de energía en la región NIR mediante los mecanismos descritos anteriormente es a través de la medida de la energía reflejada. Esta energía reflejada se expresa en unidades de absorbancia (A), definidas estas como $A = \log(1/R)$ donde R son los valores de reflectancia o cociente entre la radiación reflejada por la muestra y la radiación incidente sobre la muestra (Williams y

Norris, 1987; Burns y Ciurczak, 1992; Osborne *et al.*, 1993; ASTM, 1994, Bertrand y Dufour, 2000).

La tecnología NIRS se basa en la conocida ley de Lambert-Beer de forma que el valor de absorbancia resultante de una muestra que contenga una sustancia con capacidad de absorber radiación NIR es directamente proporcional a la concentración de la sustancia que absorba dicha radiación (Osborne y Fearn, 1986; Williams y Norris, 1987; Burns y Ciurczak, 1992 y Osborne *et al.*, 1993; Shenk y Westerhaus, 1995a; Bertrand y Dufour, 2000). Por ello, los valores de absorbancia obtenidos de una muestra se pueden relacionar con la concentración de especies existentes en la muestra, siempre y cuando absorban radiación NIR.

La representación de los valores de absorbancia a las diferentes longitudes de onda del rango NIR da lugar a una curva denominada espectro NIR el cual es el resultado de los diferentes sobretonos, bandas de combinación y absorciones electrónicas de radiación de los grupos funcionales presentes en las moléculas de la muestra (Burns y Ciurczak, 1992; Shenk y Westerhaus, 1995a; Bertrand y Dufour, 2000).

La principal consecuencia de los diferentes mecanismos anteriormente descritos es la gran complejidad, en cuanto a la interpretación espectral de las bandas de absorción, de los espectros NIR, comparados con los espectros de absorción del infrarrojo medio donde los picos o bandas obtenidas pueden ser interpretados directamente mediante su comparación con librerías espectrales de compuestos puros.

Además, para añadir más complejidad, tanto las características físicas del instrumento (compartimento para el análisis de la muestra, geometría óptica del propio instrumento, etc.) como las de la muestra (tamaño, forma y distribución de las partículas, índices de refracción, etc.) influyen en el espectro NIR de una determinada muestra. Así, en la mayor parte de las muestras es casi imposible obtener un espectro NIR sin que se produzca una dispersión de la radiación incidente, indeseable en esta técnica analítica (Burns y Ciurczak, 1992; Bertrand, 1993; Bertrand y Dufour, 2000).

Sin embargo, a pesar de que los espectros NIR resultantes son muy complejos de interpretar, el hecho de que varios enlaces moleculares estén involucrados en los diferentes tipos de absorciones de la radiación NIR anteriormente descritos,

significa que dichas absorciones pueden ser utilizadas para aportar información analítica de los enlaces moleculares o de grupos funcionales específicos (Williams y Norris, 1987; Burns y Ciurczak, 1992; Shenk y Westerhaus, 1995a; Næs *et al.*, 2002). Para poder extraer dicha información de los espectros NIR se hace necesario el uso de métodos de análisis multivariante, como se indicará posteriormente (Burns y Ciurczak, 1992; Osborne *et al.*, 1993; Næs *et al.*, 2002).

2.2.2 INSTRUMENTACIÓN Y MODOS DE ANÁLISIS

Un instrumento NIRS consta de una fuente de energía radiante, un sistema de rendijas para limitar la radiación a una banda estrecha, un sistema de lentes para enfocar la radiación sobre el sistema de dispersión de longitudes de onda o sobre la muestra, un dispositivo para la dispersión de las longitudes de onda, un compartimento para la presentación de la muestra al instrumento y unos detectores que convierten la energía radiante en una señal eléctrica. Tras la incidencia de la radiación sobre la muestra, la radiación reflejada o transmitida a través de dicha muestra es recogida por los detectores que la convierten en una señal eléctrica. Dicha señal es amplificada y convertida en una señal digital mediante un convertidor analógico/digital. Esta señal digital es transmitida a un ordenador para su almacenamiento o procesado (Shenk y Westerhaus, 1995a; Bertrand y Dufour, 2000).

El desarrollo de la instrumentación NIRS, desde finales de los años 60 hasta la actualidad, ha sido muy importante (Stark, 1996; Davies, 1999; Barton, 2002). Los cambios más relevantes que se han producido en los instrumentos NIRS serían aquellos relacionados con la mejora en la relación señal/ruido, en la precisión en la longitud de onda, en la disponibilidad de diferentes sistemas de interacción de la radiación con la muestra, en la existencia de una gran variedad de accesorios (cápsulas y cubetas de diferente forma, diferentes tipos de fibras ópticas para su utilización “in line”, “on line” y “at line”, etc.), en la mejora de los programas informáticos para la obtención y tratamiento de los espectros NIR y en la posibilidad de clonación entre diferentes instrumentos para la transferencia de ecuaciones o modelos quimiométricos entre ellos (Garrido *et al.*, 1996).

A modo de resumen indicar que, en función del tipo de dispositivo empleado para la obtención de un conjunto de radiaciones monocromáticas, se pueden diferenciar dos grandes grupos de instrumentos NIRS: equipos de filtros e instrumentos monocromadores (Davies, 1999; Barton, 2002). Ambos dispositivos se diferencian por el número de longitudes de onda que son capaces de obtener a partir de la radiación emitida por la fuente de energía radiante.

Así, mediante un filtro únicamente se puede llegar a obtener un pequeño rango de longitudes de onda (p.ej.: 10 a 20 nm de amplitud) a partir de la radiación emitida por la lámpara. Los instrumentos que poseen este dispositivo de selección de longitudes de onda vienen equipados con un sistema compuesto de un determinado número de filtros (entre 5 y 20), de forma que puedan cubrir un amplio intervalo de medida espectral en la región NIR. Los primeros equipos, comercializados en la década de los años 60 y 70 por Neotec, Dickey-John e Infrared Engineering, disponían de este sistema de selección de longitudes de onda (Davies, 1999; Barton, 2002).

Sin embargo, mediante un sistema monocromador, se puede conseguir un rango continuo de longitudes de onda a lo largo de toda la región NIRS. En la actualidad, son los sistemas más utilizados ya que proporcionan una mayor información espectral que los equipos con sistemas de filtros.

Los primeros instrumentos monocromadores, comercializados en la década de los 80 entre otras firmas por Neotec, Technicon, Tecator y Varian, disponían como sistema monocromador una red de difracción móvil (denominado *grating*). A partir de la década de los 90, comienzan a desarrollarse instrumentos equipados con nuevos sistemas monocromadores como son los diodos emisores de luz infrarroja (IRED), los interferómetros de transformadas de Fourier, los filtros ópticos-acústicos modulables (AOTF) y aquellos equipados con red de diodos en fila. Este tipo de instrumentos están comercializados hoy en día por firmas como Foss, Bran+Luebbe, Perten, Büchi, Perkin-Elmer, Zeiss o Zeltex (Davies, 1999; Barton, 2002).

Si bien en los inicios del desarrollo de la tecnología NIRS, la mayor parte de los trabajos se realizaron empleando la reflectancia como forma de interacción entre la radiación y la muestra, hoy existen diferentes configuraciones instrumentales que realizan otras modalidades de análisis (Figura 2.2).

Figura 2.2. Esquema de modalidades de análisis NIRS de reflectancia, transmitancia y doble transmitancia (adaptada de Shenk y Westerhaus, 1995a).

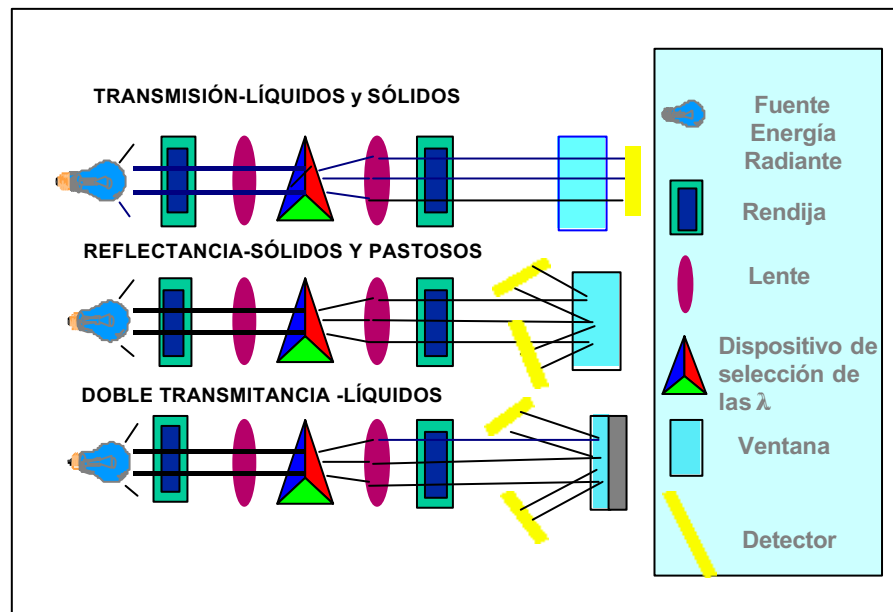
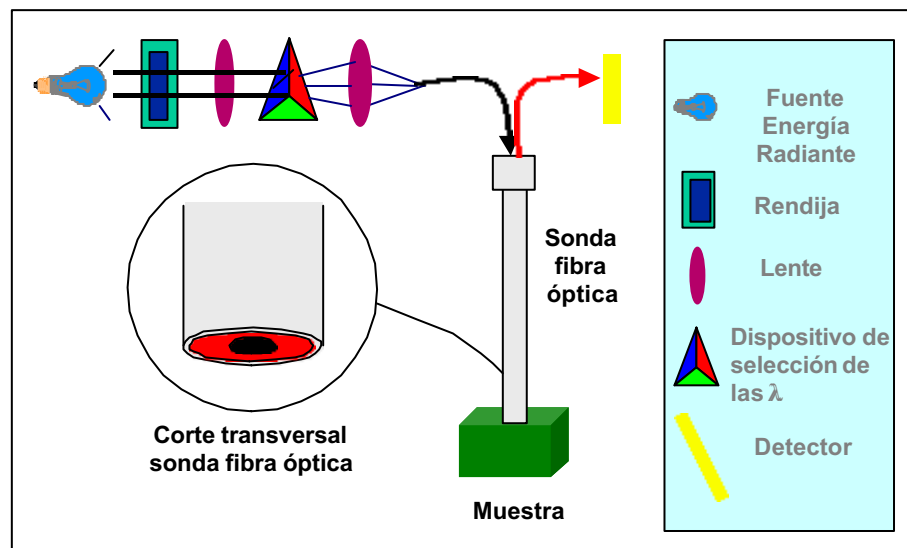


Figura 2.3. Esquema de la modalidad de análisis NIRS de interactancia-reflectancia (adaptado de Slaughter, 1995)



La más extendida junto con la reflectancia es la transmitancia. En dicha modalidad, la radiación atraviesa la muestra, de forma que se cuantifica la energía NIR absorbida por los enlaces moleculares durante su transmisión a través de la muestra. Con ello, los valores de absorbancia (A) se definen como $A = \log(1/T)$ donde T son los valores de transmitancia o cociente entre la radiación transmitida por la muestra y la radiación incidente sobre la muestra (Burns y Ciurczak, 1992; ASTM, 1994; Shenk y Westerhaus, 1995a).

Junto a estas modalidades de análisis, hoy en día se ha extendido el uso de otras modalidades de análisis que permiten formas de interacción de la radiación con la muestra más versátiles, como la DESIR (Alfaro y Meurens, 1989) o la transfectancia, ambas de gran interés en productos líquidos y pastosos, o la denominada interactancia-reflectancia, realizada mediante el uso de sondas de fibra óptica (Shenk y Westerhaus, 1995a).

La modalidad de transfectancia, también denominada doble transmitancia, fue uno de los modos de análisis introducidos por Bran+Luebbe en sus instrumentos monocromadores en la década de los 90. En esencia, esta modalidad de análisis integra una cápsula o cubeta portamuestras que permite la transmisión de la radiación NIR a lo largo de toda la muestra si bien, al final de dicha cápsula existe una superficie especular (normalmente aluminio u oro) totalmente reflectante que provoca la reflexión de la radiación, de forma que se transmite de nuevo a lo largo de toda la muestra, incidiendo sobre los detectores (Osborne *et al.*, 1993). Probablemente, debido a que esta modalidad de análisis integra un fenómeno de transmisión, por duplicado, junto con una reflexión de la radiación NIR, es por lo que posteriormente Foss NIRSystems comenzó a utilizar el término anglosajón de *folded transmission*, traducido como doble transmisión.

Esta modalidad de análisis está siendo empleada para el análisis NIRS de productos pastosos o líquidos, debido a su mayor facilidad de llenado y limpieza de cápsulas frente a otras modalidades de análisis. Así, analizar este tipo de muestras por otros modos de análisis resulta más complicado, siendo necesario, por ejemplo, el uso de filtros de fibra de vidrio empapados de la muestra líquida para su análisis mediante reflectancia, o realizar una difícil limpieza a fondo en cubetas de pequeño paso óptico mediante el análisis por transmitancia.

La modalidad de interactancia-reflectancia (Figura 2.3) supone un cambio sustancial frente a las modalidades de análisis descritas anteriormente, dado que, para el análisis de las muestras no es necesario el llenado de cápsulas o cubetas portamuestras. Así, es posible obtener los espectros de las muestras directamente sobre ellas, transmitiéndose a continuación la señal al instrumento mediante una sonda de fibra óptica (Osborne *et al.*, 1993).

En concreto, en el caso de la sonda utilizada en esta tesis, esta posee una forma cilíndrica con una longitud de 40 cm y un diámetro de 1 cm. En su interior se incluyen dos haces de fibra óptica concéntricos de forma que la radiación inciden sobre la muestra a través de la fibra óptica interna mientras que, tras interactuar con ella, es reflejada volviendo al instrumento la señal a través de la fibra óptica más externa. Dado que el instrumento NIR mide la radiación reflejada tras el pleno contacto entre la sonda y la muestra, esta modalidad de análisis se denominó interactancia-reflectancia (Shenk y Westerhaus, 1995a).

Se ha de tener presente que la presencia de la fibra óptica supone una disminución de la relación señal/ruido, provocando con ello una disminución del rango espectral NIR. En concreto, el intervalo de medida en este tipo de sondas suele considerarse, como máximo, hasta 2200 nm, dado que a longitudes de onda superiores, el ruido es muy alto.

2.2.3 QUIMIOMETRÍA Y SU APLICACIÓN EN LA TECNOLOGÍA NIRS

La Quimiometría, término introducido por Wold (1972), fue definida como la disciplina de la química analítica que mediante el uso de métodos estadísticos, matemáticos y otros de lógica formal, diseña o selecciona procedimientos y experiencias óptimos de medida y/o trata de obtener la máxima información química relevante a partir de datos químicos. Después de 30 años, la quimiometría se ha convertido en una aplicación esencial tanto dentro de la química como en otras ciencias (Massart *et al.*, 1997; Wold y Sjöström, 1998).

Como se ha indicado en el apartado 2.2.1, la gran superposición de las bandas de absorción de los diferentes constituyentes presentes en la muestra, la gran cantidad de datos existentes, la elevada información redundante (colinealidad) provocan una baja selectividad de los datos espectrales NIR. Cuando todo ello concurre, es difícil poder llegar a establecer una ecuación, mediante la ley de Lambert-Beer, para estimar una propiedad físico-química de una muestra con el valor de absorbancia a una única longitud de onda. Por ello, es muy poco útil en la tecnología NIRS la aproximación

univariante tradicionalmente empleada en el resto de técnicas espectroscópicas (Burns y Ciurczak, 1992; Osborne *et al.*, 1993; Næs *et al.*, 2002).

Además de dichos factores, la extracción de información relevante de los espectros NIR se complica debido a la presencia de numerosas fuentes de error que afectan al dato espectral. Dichas fuentes de error han sido descritas en detalle por Williams (1987), el cual agrupa las mismas en tres grupos importantes, relacionadas con el instrumento (relación señal/ruido, exactitud y precisión en la longitud de onda, linealidad de la señal, tipo y tamaño de cápsula etc.), relacionadas con la muestra (homogeneidad, densidad, textura, granulometría, estabilidad ante fluctuaciones en temperatura, etc.) y los operacionales (preparación de muestra, errores del analista, estrategias y procedimientos estadísticos empleados, etc.).

Todo lo anteriormente expuesto provoca que sea necesario emplear herramientas quimiométricas complejas en el tratamiento de los datos espectrales con el fin de obtener la información relevante (Otto, 1998). Entre dichas herramientas quimiométricas a aplicar en el análisis NIRS destacan los métodos de análisis multivariante, consideradas por Wold y Sjöström (1998) como uno de los grupos de métodos más empleados en quimiometría.

El análisis multivariante podría definirse como los diferentes métodos, ya sean estadísticos, matemáticos o gráficos, empleados en el análisis de datos siempre y cuando se consideren para dicho análisis varias variables de forma simultánea (Martens y Næs, 1989; Martens y Martens, 2000; Næs *et al.*, 2002). En el caso de la tecnología NIRS, el análisis multivariante se empleará para estimar cualquier propiedad de la muestra (físico-química, fundamentalmente) a partir de múltiples variables espectrales (valores de absorbancia a diferentes longitudes de onda).

Estas herramientas se consideran fundamentales hoy día en la obtención de información relevante a partir de los datos espectrales, siendo el desarrollo y la evolución de la tecnología NIRS paralelos a los avances registrados en dichos campos y particularmente en el terreno de la quimiometría (Shenk y Westerhaus, 1995a; Kaffka y Giamarti, 1995; Geladi, 1996).

Un aspecto que influye de forma importante en la baja selectividad de la información espectral NIR es la dispersión de la radiación incidente, también

denominado *efecto scatter* (Burns y Ciurczak, 1992; Osborne *et al.*, 1993; Bertrand y Dufour, 2000). Este hecho, debido fundamentalmente a fenómenos físicos como la textura, el tamaño y la geometría de las partículas que conforman la muestra (Iwamoto, 1991; Shenk *et al.*, 1992) y a cambios en el índice de refracción del material que interacciona con la radiación, provoca numerosas variaciones no deseadas en los datos espectrales NIR (Barnes *et al.*, 1989; Dhanoa *et al.*, 1994; Fernández y Garrido, 1999; Næs *et al.*, 2002).

Para separar la información existente en el espectro NIR debido a estas variaciones de origen físico, de aquellas variaciones que aporten información meramente química, se han desarrollado numerosos métodos quimiométricos. Dado que el origen de todo el problema es la presencia del *efecto scatter* o de radiación dispersa, a los métodos se les suele agrupar bajo el nombre de tratamientos de corrección de *scatter* (Bertrand, 1993; Bertrand y Dufour, 2000).

Algunos de estos tratamientos, según lo expuesto por Puigdomenech (1998) y Næs *et al.* (2002), serían los denominados MSC o *multiplicative scatter correction*, SNV o *standard normal variate*, DTR o *detrending* y OSC u *orthogonal signal correction*. Los tratamientos incluidos con más frecuencia en los diferentes programas de análisis multivariante aplicados a datos espectrales NIR son MSC, SNV y DTR.

Otra transformación matemática, empleada previamente a la aplicación de cualquier método de análisis multivariante sobre los datos NIRS, es el tratamiento de derivadas. Dicho tratamiento es empleado como herramienta para disminuir dos de los problemas básicos en la espectroscopía NIR como son la superposición de los picos de absorción y las variaciones de la línea base de los espectros (Shenk y Westerhaus, 1995a). Además, según Næs *et al.* (2002) también es un método muy empleado como tratamiento de corrección de radiación dispersa en los datos espectrales NIR.

Existen diferentes métodos de cálculo de las derivadas en los espectros NIR, como son el método de Savitzky-Golay, el de las transformadas de Fourier y el de las diferencias finitas (Williams y Norris, 1987; Burns y Ciurczak, 1992 y Osborne *et al.*, 1993). En este último, uno de los más utilizados en espectroscopía NIR, la derivada se asocia a la diferencia, a lo largo de todo el espectro, de absorbancias de las longitudes de ondas vecinas (Shenk y Westerhaus, 1995a).

Los datos espectrales obtenidos de un instrumento NIRS, sometidos o no a algún tipo de tratamiento de los descritos previamente, pueden ser empleados tanto en análisis cuantitativo como en cualitativo. Así, los datos de absorbancia en la región NIR pueden ser empleados para estimar la concentración de diferentes analitos presentes en una muestra, agrupándose dichas actuaciones bajo el nombre de análisis cuantitativo.

De igual forma, los datos espectroscópicos NIR por sí mismos y sin necesidad de realizar ningún tipo de análisis de referencia, pueden emplearse para estimar una variable categórica o de grupo (tipo, variedad, origen, etc.). En este caso, las actuaciones orientadas en este sentido se encuadran dentro del denominado análisis cualitativo.

2.2.3.1 *Análisis cuantitativo*

Como se ha indicado previamente, en el análisis cuantitativo, el más extendido dentro de las aplicaciones de la tecnología NIRS, se estima la concentración de un analito en una determinada muestra. Para ello se ha de establecer la relación (ecuación matemática) existente entre los valores de absorbancia en la región NIR y la concentración de dicho analito, determinada mediante un método de referencia, en muestras de composición conocida.

Este proceso se conoce con el nombre de etapa de calibración, a la relación matemática desarrollada se le denomina ecuación de calibración y el grupo de muestras de composición se designa como colectivo de calibración. Una vez establecida la ecuación de calibración, ésta permite predecir el contenido de otras muestras similares en características a las incluidas en el colectivo de calibración (Williams y Sobering, 1996; Shenk y Westerhaus, 1995b y 1996).

Conviene indicar que el proceso de obtención de ecuaciones de calibración, desde este punto de vista simplificado, puede parecer sencillo. Sin embargo, el desarrollo de ecuaciones de calibración NIRS globales o universales y robustas, es decir, ecuaciones que puedan ser de aplicación en la práctica totalidad de las muestras de una población y asimismo que mantengan su exactitud y precisión a lo largo del tiempo, requiere del seguimiento de unas estrategias concretas, algunas de las cuales se indicarán a continuación (Shenk y Westerhaus, 1995b y 1996).

Una etapa de gran importancia en el desarrollo de una ecuación de calibración NIRS es la selección de muestras que van a constituir el colectivo de calibración. En general, mientras mayor sea el número de muestras que constituyan el colectivo de calibración, mayor capacidad de predicción presentará una ecuación. Sin embargo, dado que dichas muestras han de ser analizadas mediante el método de referencia, el cual suele ser lento y costoso, el número de dichas muestras a emplear suele ser limitado. Además, ha quedado demostrado que, además de su número, un aspecto muy importante es la variabilidad, tanto espectral como química que cubran dichas muestras (Shenk y Westerhaus, 1991a, 1995b y 1996; Næs e Isaksson, 1989; Isaksson y Næs, 1990).

Si lo que se pretende es desarrollar una ecuación NIRS global, las muestras del colectivo de dicha calibración han de ser representativas del producto a analizar, abarcar la variabilidad existente, en cuanto a composición se refiere, de dichas muestras y contener todas las fuentes de variación que es posible esperar en el análisis de muestras futuras (variabilidad debida a ciclos de producción, origen de las muestras, etc.). A los colectivos de muestras que cumplen estos objetivos se les suele denominar, dentro de la terminología NIRS, librerías espectrales universales o globales de un producto (Shenk y Westerhaus, 1991a y 1995b).

Hoy en día se han propuesto numerosos métodos de selección de muestras para constituir el colectivo de calibración (Næs e Isaksson, 1989; Isaksson y Næs, 1990) e incluso, en algunos de los programas de tratamiento quimiométrico de espectros NIR, se suelen incorporar herramientas para la construcción, estructuración y selección de muestras representativas de una librería espectral NIR (Esbensen *et al.*, 1994; Shenk y Westerhaus, 1995b y 1996). La mayoría de dichas herramientas se basan en la realización de un análisis de componentes principales (ACP), seguido del cálculo de distancias entre muestras (espectros) en un espacio n-dimensional, a través generalmente de la distancia de Mahalanobis (Shenk y Westerhaus, 1995b, 1996).

El análisis de componentes principales se define en la norma ASTM (1990) como el procedimiento matemático empleado para transformar conjuntos de datos en una nuevas variables ortogonales, denominadas componentes principales, que son combinación lineal de los datos originales (en nuestro caso datos espectrales NIR), para cualquier nivel de exactitud requerido. Cada una de estas nuevas componentes

principales (CPs) calculadas intenta explicar la máxima cantidad posible de la varianza residual existente en el colectivo de datos originales. Existen numerosos y criterios y métodos estadísticos para seleccionar el número de componentes principales óptimo, a partir del cual no se aporta información adicional al modelo construido, es decir, no existe diferencia entre el valor de la varianza residual de dos componentes principales consecutivas.

El hecho de que las componentes principales sean ortogonales significa que dichas variables son linealmente independientes. Este aspecto resulta de interés para análisis NIRS ya que, en este tipo de espectroscopía se suele trabajar con un bajo número de muestras comparado con el número de variables espectrales generadas por los actuales instrumentos NIRS. Además, como ya se indicó previamente, los datos espectrales NIR presentan una gran cantidad de información redundante (colinealidad) la cual es sintetizada al emplear calcular dichas componentes principales.

Por este motivo, al ACP, además de ser una herramienta para la selección de muestras del colectivo de calibración, es una técnica de gran utilidad en el análisis para, entre otras aplicaciones, transformar los datos espectrales originales (valores de absorbancia) en nuevas variables (componentes principales) ortogonales, eliminando con ello la colinealidad (Martens y Næs, 1989; Burns y Ciurczak, 1992; McClure, 1992; Bertrand, 1993; Osborne *et al.*, 1993; Bertrand y Dufour, 2000; Martens y Martens, 2000; Næs *et al.*, 2002).

Otra de las posibilidades consideradas en análisis NIRS para evitar la colinealidad ha sido el emplear un pequeño número de longitudes de onda como variables espectrales. Esta opción, tradicionalmente empleada en las primeras etapas del desarrollo de esta técnica en instrumentos de filtros, no es tan utilizada hoy en día.

Una vez seleccionadas las muestras de calibración, analizadas estas mediante el método de referencia y NIRS y seleccionadas las variables espectrales a considerar (longitudes de onda seleccionadas, componentes principales, etc.), la siguiente etapa a considerar es el desarrollo en sí del modelo matemático que relacione ambos tipos de datos.

Para ello existen diferentes métodos de regresión (Martens y Næs, 1989; Burns y Ciurczak, 1992; Puigdomenech, 1998; Næs *et al.*, 2002), siendo los más

utilizados en las aplicaciones cuantitativas NIRS la regresión lineal múltiple (RLM ó MLR), la regresión por componentes principales (RCP ó PCR) y la regresión mediante mínimos cuadrados parciales (RMCP ó PLS).

La regresión RLM, en sus diferentes variantes (Shenk y Westerhaus, 1995b, 1996; Puigdomenech, 1998), lleva a cabo una regresión a partir de un pequeño número de variables espectrales seleccionadas de entre todo el espectro NIR. Dichas longitudes de onda, seleccionadas mediante un test F , son aquellas que presentan una mayor correlación entre el valor de absorbancia a una longitud de onda determinada y el dato de composición química.

Los métodos de regresión RCP y RMCP se asemejan en sus fundamentos ya que mediante ambos métodos se pretende generar unas nuevas variables, combinación lineal de los datos espectrales originales y emplear estas nuevas variables en el desarrollo de la ecuación de regresión (Martens y Næs, 1989; Shenk y Westerhaus, 1995b; Martens y Martens, 2000). Así, únicamente es considerada en el desarrollo de la regresión aquella información presente en los datos espectrales realmente relevante.

La principal diferencia entre ambos se basa en el tipo de combinaciones lineales que construyen a partir de los datos originales. Mientras la regresión en componentes principales (RCP) emplea los valores de las componentes principales, la regresión por mínimos cuadrados parciales (RMCP) construye unos factores RMCP. Estos factores RMCP se seleccionan de forma tal., que expliquen el mayor porcentaje de variabilidad, no solo en los datos espectroscópicos, como en el ACP, si no asimismo la mayor variabilidad en los datos de composición química. (Martens y Næs, 1989; Shenk y Westerhaus, 1995 a, Næs *et al.*, 2002). Dado que dichos factores RMCP son ortogonales, la regresión RMCP también se podría emplear como método de reducción del fenómeno de colinealidad (McClure, 1992; Bertrand y Dufour, 2000).

Los métodos de regresión indicados anteriormente permiten obtener una relación lineal entre los datos espectrales (o sus combinaciones lineales) y los de composición. Sin embargo, en algunos casos, no es posible encontrar una respuesta lineal por lo que es necesario recurrir a métodos de regresión no lineales. Entre dichos métodos destaca la regresión mediante redes neuronales y sus diferentes variantes

(Martens y Næs, 1989; Shenk y Westerhaus, 1996; Bertrand y Dufour, 2000; Næs *et al.*, 2002).

Durante el desarrollo de una ecuación de calibración NIRS, siempre se observa la presencia de determinadas muestras que, por diversos motivos, presentan un comportamiento diferente al resto de muestras presentes en el colectivo de calibración. A estas muestras se les denomina anómalas (*outliers*). La anomalía puede ser causada tanto por la información espectroscópica como por la química. La detección, interpretación y posible eliminación de estas muestras anómalas es una etapa crítica en el desarrollo de una ecuación de calibración, debido a la gran influencia que provoca su presencia en los resultados del modelo.

Por ello, existe una amplia bibliografía acerca de los tipos de anómalos existentes, cuales son las consecuencias de su presencia en un modelo de calibración, cuales son las herramientas existentes para poder detectarlos y cómo tratarlos después de su detección (Martens y Næs, 1989; Mark y Workman, 1991; Barnett y Lewis, 1994; Shenk y Westerhaus, 1995b; Martens y Martens, 2000; Næs *et al.*, 2002).

Dentro de los anómalos hay dos grandes grupos. En un primer lugar aquellas muestras que presentan unos datos espectrales (ya sea en una única longitud de onda, en varias o en todo el espectro) significativamente diferentes del resto de muestras. A estos espectros o datos espectrales también se les denomina anómalos espectrales. En un segundo lugar, estarían aquellas muestras que presentan diferencias significativas del resto del colectivo de calibración en los datos de composición química, ya sea en la aportada por el método de referencia o bien en la estimada mediante la ecuación de calibración. A este tipo de muestras también se les denomina anómalos químicos.

Para la detección de anómalos espectrales, entre los métodos más empleados se encuentran el cálculo de distancias en espacios n-dimensionales, como puede ser la distancia de Mahalanobis o el estadístico *leverage*, y el cálculo de residuales en los datos espectrales, es decir, espectros o datos espectrales que presenten una parte importante de su variabilidad sin explicar por el modelo generado (Martens y Næs, 1989; Shenk y Westerhaus, 1995a y b).

Los métodos quimiométricos a emplear para la generación de dichos modelos y, por tanto, para el cálculo de los estadísticos propuestos anteriormente y la detección de anomalos espectrales, pueden ser tanto un análisis de componentes principales como una regresión propiamente dicha mediante componentes principales (RCP) o factores RMCP.

Ambos tipos de estadísticos (distancias en espacios n-dimensionales y residuales) pueden ser también empleadas en la detección de anomalos químicos. Sin embargo, la más extendida es el cálculo de residuales, en este caso, en los datos de composición. Dichos residuales químicos se definen como aquellas observaciones donde se aprecian unas diferencias significativamente diferentes, evaluadas normalmente mediante un test T , entre los valores de composición aportados por el método de referencia y los estimados mediante el modelo de regresión.

Una vez detectada la presencia de anomalos químicos y espectrales, estas muestras han de ser interpretadas mediante la búsqueda de causas que pudieran provocar dichas diferencias, repitiendo en la medida de lo posible su análisis espectral o de referencia (Shenk y Westerhaus, 1995b; Martens y Næs, 1989; Næs *et al.*, 2002).

Después de haber desarrollado una ecuación de calibración mediante alguno de los métodos de regresión existentes, es esencial determinar su capacidad de predicción de muestras que no hayan intervenido en la calibración. Este proceso se denomina etapa de validación (o validación externa) mientras que las muestras empleadas en dicha etapa se designan como colectivo de validación.

Una variante de la validación descrita anteriormente es la denominada validación cruzada. Esta técnica, descrita por Stone (1974), consiste en una validación basada únicamente en los datos empleados en la etapa de calibración. Para ello, se eliminan de forma sucesiva muestras del colectivo de calibración que son empleadas para validar la ecuación desarrollada a partir del resto de muestras del colectivo de calibración que no han sido eliminadas. El procedimiento se repite hasta lograr que todas las muestras hayan sido utilizadas una vez para realizar la validación. Con ello se logra obtener una estimación de la capacidad de predicción de una ecuación más real que la obtenida durante el desarrollo del modelo de calibración.

Desde el punto de vista estadístico, la etapa de validación consiste en la comparación, mediante diferentes criterios estadísticos, del valor verdadero y el valor estimado (Massart *et al.*, 1988). Particularizando en la tecnología NIRS, se considerará como valor verdadero aquel determinado mediante el método de referencia mientras que el valor estimado será aquel predicho por la ecuación de calibración NIRS.

Los criterios estadísticos más importantes para evaluar el error o la capacidad de predicción de un procedimiento analítico son la exactitud y precisión. (Massart *et al.*, 1988). Se han propuesto diferentes definiciones de exactitud y precisión si bien, aquella más extendida es la de considerar la exactitud como la diferencia entre el valor verdadero y el valor estimado para un conjunto de muestras mientras que la precisión se refiere a las diferencias existentes entre medidas repetidas de unas mismas muestras (Massart *et al.*, 1997; Næs *et al.*, 2002).

Un estimador acerca de la capacidad de predicción de un modelo es el propio error de calibración del modelo desarrollado. Este error típico de calibración (*ETC*) se define como la desviación típica de los residuales (diferencia entre el valor aportado por el método de referencia y el valor estimado por la ecuación) para el colectivo de calibración. Sin embargo, dicho estadístico al ser únicamente un estimador del error del modelo quimiométrico empleado, suele sobrestimar la capacidad real de predicción de una ecuación (Shenk y Westerhaus, 1995b; Mark y Workman, 1991; Martens y Næs, 1989; Næs *et al.*, 2002).

El estadístico más empleado en las diferentes aplicaciones cuantitativas NIRS para estimar la capacidad de predicción de una ecuación de calibración es el error típico de predicción (*ETP*), siendo este la desviación típica de las diferencias, para un colectivo de validación, entre el valor determinado mediante el método de referencia y valor estimado mediante el análisis NIRS. En el caso de emplear el procedimiento de validación cruzada, el estimador, similar al *ETP*, se denomina error típico de validación cruzada (*ETVC*).

El estimador *ETP* puede ser dividido en dos tipos de errores diferentes, como son los errores aleatorios o no explicados y los errores sistemáticos o sesgo (Windham *et al.*, 1989; Massart *et al.*, 1988 y 1997). Por ello, el estadístico *ETP* se puede dividir a su vez en otros dos estadísticos que reflejen, de forma independiente,

dichos errores sistemáticos y aleatorios. Ello es posible mediante la conocida fórmula de la partición de la varianza del error (Martens y Næs, 1989; Næs *et al.*, 2002) descrita a continuación:

$$ETP^2 \cong sesgo^2 + ETP(C)^2$$

donde el estadístico *sesgo* informará acerca de los errores de tipo sistemático mientras que el estadístico *ETP(C)* informará sobre los errores aleatorios. De acuerdo con Næs *et al.* (2002), el estadístico *ETP* evalúa la exactitud de una ecuación mientras que el estadístico *ETP(C)* evalúa su precisión.

Es importante resaltar que el estadístico *ETP(C)* definido previamente es el empleado por algunos autores (Williams y Norris, 1987; Shenk y Westerhaus, 1995a) y programas quimiométricos (p.ej.: ISI). Sin embargo, otros autores (Martens y Næs, 1989; Næs *et al.*, 2002) o programas (p.ej.: Unscrambler) denominan a dicho estadístico como *ETP*.

Junto a lo anterior, conviene destacar la importancia de la calidad del dato analítico de referencia en la aplicación cuantitativa NIRS. Así, según indica Fearn (1986), el error de predicción también puede expresarse en función del error del método de referencia a partir de la expresión:

$$ETP^2 = ETL^2 + ET_{NIR}^2 + ET_{Modelo}^2$$

donde el estadístico *ETL* es el error típico del método de referencia, el estadístico *ET_{NIR}* es el error típico debido a la medida espectral NIR, mientras que el estadístico *ET_{Modelo}* es el error introducido por el método quimiométrico. Las mejoras en la instrumentación y en los programas de tratamiento quimiométrico han permitido reducir sensiblemente el error debido a la medida espectroscópica NIR y el error del modelo quimiométrico, pero el error de predicción de la ecuación desarrollada, aún viene condicionado en gran medida, por el error del método de referencia, el cual es ajeno a la propia tecnología NIRS.

Se ha de tener en cuenta que, durante la etapa de predicción también se pueden presentar muestras anómalas, tanto químicas como espectrales. Estas muestras no van a afectar a la ecuación generada, si bien se ha de tener en cuenta que la

predicción en dichas muestras suele ser errónea. En predicción, dado que no siempre se va a disponer de muestras con valores de composición determinados mediante el método de referencia, la mayor parte de la detección de anómalos se centra en la búsqueda de anómalos espectrales. Para ello se emplean las mismas herramientas propuestas anteriormente, fundamentalmente distancias como la de Mahalanobis o *leverage*.

A modo de resumen, indicar que Shenk (1985), Shenk *et al.*, (1992), Esbensen *et al.* (1994), ASTM (1994, 2000), Shenk y Westerhaus (1995b) y Næs *et al.* (2002) proponen diferentes protocolos a emplear como guía a la hora de desarrollar y evaluar una aplicación cuantitativa NIRS.

2.2.3.2 *Análisis cualitativo*

El análisis cualitativo NIRS consiste en la comparación de espectros de muestras desconocidas con grupos de espectros NIR de muestras de características conocidas (Mark, 1992). Estas características o atributos, variables categóricas y no continuas, pueden ser, a modo de ejemplo, el origen geográfico, variedad, tipo de producto, tipo de alimentación, etc.

Se han desarrollado numerosos métodos quimiométricos multivariantes para poder llegar a establecer un análisis cualitativo a partir de datos químicos (Massart *et al.*, 1997), habiéndose aplicado la mayoría de ellos a datos espectrales NIR (Mark, 1992; McClure, 1992; Bertrand, 1993; Kaffka y Giamarti, 1995; Downey, 1996 y 2000; Martens y Martens, 2000; Dardenne, 2000a; Næs *et al.*, 2002). Incluso, algunos de dichos algoritmos, ya fueron incorporados en las primeras versiones de los programas informáticos para la recogida y tratamiento estadístico de espectros NIR (Mark, 1992).

Todos los métodos de clasificación o de reconocimiento de pautas se podrían agrupar en dos grandes categorías: supervisados y no supervisados. Ambos se diferencian en el nivel de información que se disponga *a priori* del agrupamiento existente en los datos (Massart *et al.*, 1997).

Así, si no se dispone de ninguna información o es escasa acerca de los grupos (categorías) existentes en las muestras analizadas, los métodos se denominarán no supervisados. El principal objetivo de estos métodos es poder identificar tendencias

al agrupamiento de las muestras disponibles, sin necesidad de disponer de ningún tipo de información adicional acerca de ellas nada más que sus datos espectrales (Bertrand y Dufour, 2000; Downey, 1996 y 2000; Næs *et al.*, 2002). Por ello, estos métodos suelen ser aplicados en las primeras etapas de la investigación o exploración de los datos. Un ejemplo de dichos métodos es el análisis cluster y sus diferentes variantes (Massart *et al.*, 1997), si bien dentro de esta categoría también se puede incluir al análisis de componentes principales (Næs *et al.*, 2002).

A diferencia de los métodos no supervisados, en los métodos supervisados, frecuentemente empleados en análisis cualitativo NIRS, si se dispone de información acerca de los grupos o categorías existentes en el colectivo de muestras. El principal objetivo de este tipo de métodos es el de generar diferentes reglas o modelos de clasificación para un número determinado de clases o grupos de muestras analizadas (Massart *et al.*, 1997; Bertrand y Dufour, 2000; Downey, 1996 y 2000; Næs *et al.*, 2002). Para ello se parte de un conjunto de muestras, denominado grupo de aprendizaje, del que se dispone tanto de los datos espectrales NIR como de la información de la variable categórica (origen, procedencia, variedad), estableciéndose a partir de ellos una función que correlacione ambos tipos de datos.

Se ha de destacar que, la principal diferencia del análisis cualitativo con el análisis cuantitativo anteriormente descrito es que, en este caso, no es preciso disponer de ningún dato químico adicional al dato espectroscópico NIR, por lo que el colectivo de aprendizaje sólo está constituido por información espectral NIR, de ahí que se pueda denominar análisis espectral *per se*.

Una vez generados los modelos o reglas de clasificación, estos son empleados para asignar nuevas muestras, de características desconocidas, a aquella categoría a la que presente mayor semejanza o similitud en términos estadísticos. Dicha similitud suele expresarse en términos de distancia en un espacio n-dimensional. Existen diferentes estadísticos para poder expresar esta distancia como son la distancia euclídea, de Mahalanobis, de Manhattan, de Minkowski, etc. (SAS, 1999; Massart *et al.*, 1997; De Maesschalck *et al.*, 2000). También puede evaluarse la similitud en términos de probabilidad de pertenencia a una categoría (Mark, 1992; Massart *et al.*, 1997).

El análisis cualitativo también puede ser utilizado como herramienta en la interpretación de diferencias entre grupos de muestras, en la localización de bandas de absorción características de grupos funcionales, en la detección de zonas de máxima diferencia entre dos productos, etc. (Williams y Norris, 1987b; Downey, 1996 y 2000; Martens y Martens, 2000; Næs *et al.*, 2002).

Entre los métodos de clasificación supervisados, los de mayor aplicación a datos espectrales NIR son el análisis discriminante lineal (ADL) y cuadrático (ADC), análisis discriminante canónico (o análisis discriminante lineal de Fisher), análisis SIMCA (Soft Independent Modelling of Class Analogies), análisis mediante redes neuronales artificiales, análisis KNN (*K nearest neighbours*) y análisis discriminante mediante regresión RMCP o RCP (Martens y Næs, 1989; Mark, 1992; Downey 1996; Bertrand y Dufour, 2000; Næs *et al.*, 2002).

Básicamente, en el análisis discriminante lineal, se realiza la separación entre las categorías existentes en el colectivo de aprendizaje a partir de una combinación lineal de las variables empleadas que minimice la variabilidad de las muestras de una misma clase y maximice la variabilidad entre clases. A partir de estas combinaciones lineales se generan funciones lineales, de forma que si existen n clases se calculan $n-1$ funciones discriminantes (Massart *et al.*, 1997; Næs *et al.*, 2002).

Para ello, se asume una distribución normal de las variables empleadas en el análisis y la igualdad entre las matrices de covarianzas. Como criterio de clasificación suele emplearse la distancia de Mahalanobis (Mahalanobis, 1936; De Maesschalck *et al.*, 2000), considerando para su cálculo la matriz de covarianzas ponderada entre las diferentes clases, o bien la probabilidad de pertenencia a una determinada categoría.

Una variante gráfica del análisis ADL es el análisis discriminante canónico. Presenta el mismo fundamento, es decir, busca direcciones en el espacio multivariante donde la separación entre los grupos sea máxima, definiendo a continuación una nuevas variables (variables canónicas) en dichas direcciones. El resultado más empleado para este análisis suele ser la representación bidimensional de determinadas variables canónicas, visualizando con ello la discriminación existente entre las clases del grupo de aprendizaje. Sin embargo, mediante este análisis no es posible la clasificación de

muestras no incluidas en el grupo de aprendizaje (únicamente tiene carácter exploratorio), hecho este que sí permite el análisis ADL (SAS, 1990).

En el análisis ADL se asume la igualdad entre las matrices de covarianzas de las diferentes clases, hecho que no siempre es cierto. Cuando esto ocurre, se emplea un análisis similar al análisis ADL si bien, se tiene en cuenta para los cálculos de las funciones discriminantes cada una de las matrices de covarianzas de las clases en vez de considerar la matriz de covarianzas ponderada. El resultado de dicho análisis es una función cuadrática por lo que el análisis se denomina análisis discriminante cuadrático (ADC). Igualmente, para el cálculo de la distancia entre grupos (normalmente, la distancia de Mahalanobis), se consideran las matrices de covarianzas de cada clase de forma individualizada (SAS, 1990; Mark, 1992; Næs *et al.*, 2002).

El análisis SIMCA, propuesto por Wold (1976), se basa en la realización de un análisis de componentes principales para cada una de las clases que componen el colectivo de aprendizaje. Este análisis multivariante difiere de los anteriormente descritos en que se desarrolla una descripción matemática (modelado) de cada clase, independientemente del resto de clases que componen el colectivo de aprendizaje. Desde un punto de vista geométrico, este modelado se correspondería con la construcción de un recinto cerrado para cada clase, a partir de su centroide y una dispersión de los datos frente a dicho centroide determinados mediante un análisis de componentes principales (Massart *et al.*, 1988 y 1997).

A este tipo de análisis en el que se construyen recintos cerrados para cada clase se denomina análisis de modelado, para diferenciarlo de los anteriormente descritos denominados frecuentemente como análisis de clasificación, en los que los recintos para cada clase eran abiertos (ADL, ADC y análisis discriminante canónico).

Dado que en todo análisis de modelado, y en concreto en el análisis SIMCA, se crea un modelo para cada clase, también son diferentes las posibilidades a la hora de asignar una muestra como perteneciente a las categorías establecidas. Así, en este caso se podrían dar 3 posibilidades: la muestra no pertenece a ninguna de las clases, pertenece a una de ellas o bien, puede pertenecer a 2 ó más clases. Esta última posibilidad se dará siempre que exista una región en la que solapen varias clases y la muestra a asignar se ubique en dicha región (Massart *et al.*, 1988).

En el análisis SIMCA, los criterios empleados para considerar a una muestra como perteneciente a una determinada clase son dos: distancia de la muestra al centro del modelo y distancia de la muestra al modelo. La primera de ellas informa acerca de la varianza que queda explicada por el modelo de una determinada clase cuando la muestra se proyecta sobre las componentes principales (se suele emplear la distancia de Mahalanobis). La segunda de ellas, que representa la distancia de la muestra al espacio definido por las componentes principales, informa sobre la varianza que no puede ser explicada por el modelo, denominándose *residual* o *S* (Martens y Næs, 1989; Esbensen *et al.*, 1994; Martens y Martens, 2000).

Otro tipo de análisis de modelado es el denominado análisis discriminante mediante regresión. Se basa en el hecho de considerar el análisis discriminante como una calibración cualitativa de forma que, en vez de calibrar para una variable continua, se desarrolle una calibración para la pertenencia o no a una determinada clase (Shenk y Westerhaus, 1995a; Massart *et al.*, 1997; Næs *et al.*, 2002).

Este análisis se basa en la creación de una falsa variable (una por cada clase considerada) de forma que a la pertenencia a una clase se le asigne el valor de 1 mientras que a su no pertenencia se asigne 0. Estos valores son arbitrarios de forma que se podrían elegir otros como $-1/1$ ó $1/2$ de forma que los resultados serían los mismos (Næs *et al.*, 2002). A partir de aquí, se desarrolla una ecuación de regresión que correlacione estas falsas variables con las medidas espectrales. Como algoritmo de regresión suelen emplear aquel basado en las componentes principales (RCP) o por mínimos cuadrados parciales (ya sea RMCP1 si únicamente hay dos clases ó RMCP2 cuando hay más de dos clases).

Mediante este análisis, una muestra se considerará como perteneciente a una clase si supera un determinado valor límite. Así, si las falsas variables toman los valores de $1/2$, el valor límite será 1,5, de forma que toda muestra que presente para una falsa variable (es decir, clase) un valor superior a 1,5 se considera como perteneciente a ella mientras que si el valor es inferior, no pertenecerá (McElhinney *et al.*, 1999). Otros autores consideran que una muestra se asignará a aquella clase en la cual presenten el máximo valor para las falsas variables (Næs *et al.*, 2002).

Un aspecto de gran importancia a la hora de emplear cualquiera de los métodos de análisis cualitativo es el tipo de variables a emplear. Estas pueden ser, o bien un número seleccionado de longitudes de onda (Mark, 1992) o bien emplear todos los datos espectrales disponibles. Hoy en día, la mayor parte de las aplicaciones cualitativas se desarrollan partiendo de un número elevado de datos espectrales NIR, a los cuales se les ha aplicado algún tipo de tratamiento de derivadas o de corrección de la radiación difusa.

Sin embargo, dada la colinealidad existente y la presencia, en la mayoría de las ocasiones, de un menor número de muestras que variables espectrales NIR es prácticamente imposible emplear los valores de absorbancias a diferentes longitudes de onda directamente como variables en los diferentes métodos de análisis cualitativo. Por ello, del mismo modo que en las aplicaciones de análisis cuantitativo, es necesario eliminar dicha información redundante y reducir el número de variables, generalmente mediante un análisis de componentes principales o mediante las variables RMCP (Massart *et al.*, 1997; Næs *et al.*, 2002).

En la mayoría de las aplicaciones cualitativas NIRS (Downey, 1996 y 2000) las variables empleadas son los valores de las componentes principales del colectivo de aprendizaje, simplificando enormemente el cálculo de los modelos matemáticos. En aquellos casos donde el análisis cualitativo lleve implícito una etapa previa de eliminación de la colinealidad, como en el caso del análisis SIMCA o discriminante mediante regresión, existe la posibilidad de partir o bien de los datos espectrales sin transformar ($\log 1/R$) o tras una transformación previa (p.ej.: valores de componentes principales). Según Massart *et al.* (1997) ambas vías son correctas, si bien la mayoría de los autores prefieren emplear como variables los datos espectrales directamente.

La presencia de muestras anómalas es también un aspecto de gran importancia en el análisis cualitativo debido a la necesidad de constituir el grupo de aprendizaje a partir de muestras de características muy controladas y definidas. Las herramientas de detección de anomalías son similares a las empleadas en análisis cuantitativo. En general, muestras detectadas como de elevadas diferencias al resto en cuanto a los valores de distancia de Mahalanobis y/o varianza residual, tanto durante el desarrollo del modelo como en su posterior evaluación con muestras externas, han de ser consideradas como anomalías (Mark, 1992; Martens y Næs, 1989; Næs *et al.*, 2002).

La validación o evaluación de un modelo cualitativo es tan importante en análisis cualitativo como en las aplicaciones cuantitativas. En general, el estadístico a emplear en cualquier método de clasificación es el error, entendido este como el porcentaje de muestras clasificadas incorrectamente. Este error de clasificación puede estimarse, tanto mediante un proceso de validación cruzada como mediante un conjunto de muestras que no hayan intervenido en el desarrollo del modelo, entendiendo que mientras menor sea su valor, mayor capacidad de clasificación presentará el modelo (Næs *et al.*, 2002). Sin embargo, dicho estadístico no es posible aplicarlo en análisis de modelado donde la muestra no ha de pertenecer de forma forzosa a una determinada clase (análisis SIMCA o discriminante basado en regresión).

En estos casos, se suelen emplear otros estadísticos que informan acerca de los resultados del modelado por parte de las diferentes clases existentes. Unos de los más empleados son el número o porcentaje de *falsos positivos* y el de *falsos negativos*. (Massart *et al.*, 1988 y 1997).

Por *falsos negativos* se entiende el número de muestras que perteneciendo a una clase no han sido asignados a dicha clase por su correspondiente modelo. Por *falsos positivos* se entiende el número de muestras que, a pesar de no pertenecer a una determinada clase, son reconocidas como tales por el modelo de dicha clase. Mientras menor sea el valor de ambos estadísticos menor error a la hora de clasificar presentarán los modelos.

A modo de resumen, indicar que Massart *et al.* (1988) y Næs *et al.* (2002) proponen diferentes estrategias a emplear como guía a la hora de desarrollar y evaluar una aplicación de análisis cualitativo.

2.2.3.3 Clonación / estandarización de instrumentos y transferencia de calibraciones

Un factor importante para la implantación de la tecnología NIRS a nivel de la industria y laboratorios públicos y privados es el desarrollo de ecuaciones de calibración globales y robustas (Shenk y Westerhaus, 1995b) que como se ha indicado con anterioridad se denominan a aquellas ecuaciones de calibración obtenidas a partir de un gran número de muestras y que pueden ser empleadas para predecir con precisión y

exactitud la mayor parte de las muestras de una población y que mantengan su precisión y exactitud a lo largo del tiempo.

Las ecuaciones generadas teniendo en cuenta dichos factores, no sólo tienen un alto valor científico sino también un alto valor económico. Por lo tanto, sería deseable que una vez fueran obtenidas dichas ecuaciones NIRS globales y robustas, éstas pudieran ser empleadas en diferentes instrumentos localizados en el mismo laboratorio o bien en laboratorios diferentes. Para ello, es necesario que el espectro de una misma muestra obtenido en el instrumento donde fueron generadas las ecuaciones (instrumento Master) sea idéntico al espectro obtenido en aquel instrumento NIR hacia el cual se desean transferir las ecuaciones (instrumento Satélite).

Hoy en día, a pesar del avance instrumental acaecido en los últimos años, aun los espectros obtenidos en diferentes instrumentos, incluso del mismo fabricante, presentan pequeñas diferencias derivadas de la configuración o de la óptica del instrumento. Ello hace imposible la transferencia directa de ecuaciones entre instrumentos, si lo que se pretende es emplear dichas ecuaciones con la misma exactitud y precisión que en el instrumento Master.

Tradicionalmente, el método empleado para resolver el problema de la transferencia de ecuaciones era el ajuste, mediante corrección de la pendiente y/o sesgo o bien mediante ajuste de coeficientes y longitudes de onda, de las ecuaciones a cada equipo Satélite en particular. Además de tedioso, dicha solución únicamente resolvía el problema de forma transitoria, siendo además imposible el compartir los espectros entre instrumentos para realizar calibraciones comunes (Dardenne *et al.*, 1992).

Durante la 3ª Conferencia Internacional NIRS celebrada en Bélgica, fue presentado un procedimiento para llevar a cabo la corrección de las diferencias existentes entre espectros de las mismas muestras obtenidos en instrumentos diferentes. Este procedimiento, denominado estandarización o clonación de instrumentos, se basa en el uso de lo que se conoce como “set de universales”, esto es, 30 muestras de diferentes productos agrícolas (cereales, alimentos proteicos, forrajes, subproductos, etc.) las cuales se encuentran incluidas en cápsulas selladas para evitar, en la medida de lo posible, sus alteraciones (Shenk, 1990).

Posteriormente, fue propuesto por Shenk y Westerhaus (1991b) un método de estandarización simplificado para la nueva generación de instrumentos NIR. Este procedimiento simplificado únicamente emplea una muestra de un producto agrícola proteico, envasada en una cápsula de análisis sellada, para corregir las diferencias espectrales existentes entre instrumentos.

Hoy en día, se han desarrollado numerosas herramientas quimiométricas para poder realizar la estandarización o clonación entre instrumentos NIR y la posterior transferencia de ecuaciones (Bouveresse y Massart, 1996; Dardenne, 2000b; Fearn, 2001; Næs *et al.*, 2002). Entre ellos se pueden citar el método DS (*Direct standardisation*) y PDS (*Piecewise direct standardisation*) descrito por Wang *et al.* (1991), aquel que emplea OSC (Ortogonal signal correction) propuesto por Wold *et al.* (1998) así como la estandarización mediante redes neuronales (Despaigne *et al.*, 1998). Sin embargo, desde el punto de vista práctico, uno de los métodos de estandarización más utilizado sigue siendo el descrito en un principio por Shenk y Westerhaus, ya sea empleando una única muestra o varias muestras selladas.

Este método, patentado por los autores en su versión a partir de un conjunto de muestras selladas (Shenk y Westerhaus, 1989), realiza una corrección univariante del espectro completo mediante dos etapas independientes. Una primera etapa de corrección de las longitudes de onda y una segunda de corrección de los valores de absorbancias.

La corrección de longitudes de onda individuales entre los espectros del instrumento Master y el Satélite, transformados por un pretratamiento previo de primera derivada, se realiza buscando la longitud de onda del instrumento Satélite para la que sus valores de absorbancia estén más correlacionadas con cada una de las del Master. A continuación se ajusta, mediante un modelo cuadrático, a la longitud de onda con la máxima correlación y sus dos longitudes de onda más próximas. Las distintas localizaciones de los modelos cuadráticos obtenidos para cada longitud de onda del instrumento Master se consideran las longitudes de onda del instrumento Satélite que mejor se ajustan a las correspondientes longitudes de onda del Master.

A continuación se ajusta un nuevo modelo cuadrático que relaciona las longitudes de onda del instrumento Master con sus correspondientes longitudes de onda en el instrumento Satélite, obteniéndose de este modo los valores definitivos para las

longitudes de onda del instrumento Satélite correspondientes a las longitudes de onda del instrumento Master.

En la etapa de corrección de los valores de absorbancia se llevan a cabo interpolaciones para calcular las respuestas del espectro medido en el instrumento Satélite a las longitudes de onda sugeridas por el modelo cuadrático obtenido en la etapa anterior.

La corrección de las absorbancias se obtiene mediante regresión lineal de las respuestas del instrumento Satélite obtenidas por interpolación a cada longitud de onda seleccionada, sobre las respuestas del instrumento Master a la correspondiente longitud de onda. Longitud de onda por longitud de onda, se ajusta la respuesta del instrumento Satélite con los correspondientes coeficientes de regresión.

Diversos autores han demostrado la utilidad de dicho método para ajustar los espectros obtenidos en varios monocromadores NIR en diferentes productos agrícolas (Shenk, 1992; Dardenne *et al.*, 1992; Moya *et al.*, 1995; Puigdomenech *et al.*, 1995; Flinn y Saunders, 1995; Park *et al.*, 1999; Garrido *et al.*, 1999 y Osborne *et al.*, 1999).

Este algoritmo de estandarización propuesto por Shenk y Westerhaus (1990 y 1991) recomienda el uso de cápsulas selladas conteniendo muestras de productos agrícolas. Los espectros obtenidos de dichas muestras selladas debían cubrir el rango de variabilidad de la absorción espectral de los productos sobre los cuales se aplicaría el proceso de estandarización.

En principio, para llevar a cabo el proceso de clonación, se han de obtener espectros de estas muestras selladas en el instrumento donde las ecuaciones de calibración han sido generadas (instrumento Master) así como en el instrumento donde se desea implantar o transferir dichas ecuaciones (instrumento Satélite).

Mediante el uso de cápsulas selladas, se asegura que ambos instrumento obtienen idénticos espectros de la misma muestra con los que poder corregir las diferencias existentes entre los instrumentos Master y Satélite. Además, con el uso de estas cápsulas se lograría evitar, en la medida de lo posible, aquellos cambios en las

muestras debidos a variaciones de temperatura o humedad provocados durante la toma de espectros.

Sin embargo, en numerosas ocasiones no es posible el uso de cápsulas selladas, por razones prácticas o económicas. Los aspectos prácticos se deben a que, incluso con el empleo de cápsulas selladas, se pueden producir cambios físicos y/o químicos de las muestras en el interior de las cápsulas (p. ej.: muestras con gran contenido en humedad o grasa. Los motivos económicos son debidos al encarecimiento del proceso de estandarización al emplear cápsulas de alto coste (p. ej.: cápsulas de transporte rectangulares o cápsulas circulares de doble transmisión).

Hoy en día, se han propuesto sin embargo otros métodos de estandarización mediante los cuales no es necesario el uso de muestras. Este es el caso del descrito por Blank *et al.* (1996) mediante el cual se aplica una corrección MSC individualizada sobre cada uno de los datos espectrales. Este método fue desarrollado para productos de origen farmacéutico, no habiendo sido evaluado aun para productos de origen agroalimentario (Næs *et al.*, 2002).

Park *et al.* (1999) han mostrado la posibilidad de clonar instrumentos y por tanto de transferir ecuaciones de calibración NIR para análisis de ensilado fresco sin el uso de cápsulas selladas e incluso presentando diferentes superficies de muestra a los instrumentos Master y Satélite. Sin embargo, se desconoce la metodología a utilizar, particularmente en lo referente al tipo de muestra de referencia, en el caso de transferencia de ecuaciones de muestras líquidas o semilíquidas como las utilizadas en la presente Tesis Doctoral.

2.3 APLICACIONES DE LA TECNOLOGÍA NIRS EN GRASAS Y ACEITES

Las primeras referencias científicas al análisis NIRS de grasas y aceites aparecen en Osborne y Fearn (1986), Williams y Norris (1987a) y Osborne *et al.* (1993). Dichas publicaciones, junto con otras publicadas más recientemente (Riaublanc *et al.*, 2000), hacen referencia a trabajos pioneros en el análisis NIRS de grasas tales como los de Holman y Edmonson (1956), Holman *et al.* (1958, 1959) y Goddu (1957, 1958).

Estos trabajos realizados durante la década de los años 50 difieren en numerosos aspectos de los realizados hoy en día, fundamentalmente por el tipo de instrumentación empleada y por el tratamiento de los datos espectrales. Es por ello que en esta Tesis únicamente se revisarán aquellos trabajos realizados en las últimas décadas si bien, donde proceda, se sintetizarán los aspectos más relevantes de los trabajos realizados anteriormente.

En un primer apartado se revisaran aquellos trabajos relacionados con la caracterización y asignación de bandas de los espectros NIRS de grasas y aceites. El segundo apartado se enfocará hacia la revisión de los trabajos científicos en los que se haya estudiado la aplicación de la tecnología NIRS en la cuantificación de parámetros analíticos comúnmente empleados en el control de calidad de grasas y aceites. Por otra parte, en un tercer apartado de esta sección se sintetizarán todos aquellos trabajos relacionados con el uso de los datos espectroscópicos NIR para el análisis cualitativo (caracterización, evaluación del grado de pureza, discriminación o autenticación) de grasas y aceites.

2.3.1 ANÁLISIS ESPECTRAL DE ASIGNACIÓN DE BANDAS

Holman y Edmonson (1956), Holman *et al.* (1958 y 1959), Goddu (1957 y 1958), Fenton y Crisler (1959), Fearn (1982) y Wetzel (1983) entre otros, aportan una valiosa información sobre el análisis espectral NIR de grasas y aceites.

De dichos trabajos es posible concluir que los ácidos grasos presentan picos de absorción a 1.700 (primer sobretono) y 1.200 nm (segundo sobretono) debidos a la vibración (tensión) del enlace C–H, junto a bandas de combinación a 2.200 y 2.500 nm de la tensión del enlace C–H con otros modos vibracionales. Además, se detectan bandas de tensión del enlace C–H de los grupos metilo (–CH₃) a 1.690, 1.730 y 2.260 nm, mientras que los enlaces C–H del grupo metileno (–CH₂) presentan bandas a 1.740, 1.770, 2.300 y 2.340 nm.

De manera más específica, los ácidos grasos con dobles enlaces *cis* poseen diferentes bandas de absorción, entre las que destacan bandas de combinación a 2.150 y 2.190 nm, un primer sobretono a 1.680 nm y un segundo sobretono a 1.180 nm, observándose que las bandas presentes a 2.150 y 2.190 nm incrementan sus valores de absorbancia cuanto mayor es el grado de insaturación presente en el ácido graso.

Murray (1987) realizó un estudio de asignación de bandas de la región NIR en compuestos orgánicos, el cual tenía, entre otros objetivos, evaluar la influencia de diversos grupos funcionales (p.ej.: –CH) en los espectros NIR de compuestos de una misma familia (p.ej.: ácidos grasos lineales saturados e insaturados).

Dicho estudio mostraba la existencia de bandas características de absorción a 1.212, 1.725, 1.760 y 2.308 nm en los compuestos orgánicos presentes de forma mayoritaria en grasas y aceites. Las bandas existentes en la región de 2308 a 2500 nm corresponderían, en opinión de Murray (1987), a bandas de combinación de modos de tensión simétricos y asimétricos del enlace C–H, mientras que los picos presentes a 1.725 y 1.760 nm serían debidos a los primeros sobretonos de tensión del enlace C–H de los grupos metilo y metileno.

Murray (1987) observó que los espectros de los ácidos grasos C18:1, C18:2 y C18:3 mostraban la presencia de unos picos nítidos a 2.150 y 2.180, los cuales no se observaron en ácidos grasos saturados, por lo que confirmó que dicha región podría ser

de gran utilidad para determinar el grado de insaturación y por tanto para la determinación del índice de yodo de aceites y grasas.

Osborne y Fearn (1986) obtuvieron similares conclusiones si bien utilizaron en este caso los espectros de aceite de cacahuete y parafina. Las principales bandas de absorción en ambos espectros se detectaron a 1.200, 1.734, 1.765, 2.310 y 2.345 nm. Sin embargo, en el caso del aceite de cacahuete se obtuvieron unas bandas específicas a 2.140 y 2.190 nm, las cuales no se detectaron en la parafina líquida. Por ello, estos autores indicaron que dichas bandas podrían ser debidas a la absorción de la radiación NIR por parte de los dobles enlaces C=C *cis* presentes en los ácidos grasos del aceite y ausentes en parafina.

A pesar de estas diferencias existentes en los espectros NIR de ácidos grasos puros, una de las características más relevantes de los espectros de grasas y aceites es la escasa diferencia en los valores de absorbancia a lo largo de todo el espectro NIR incluso para aceites o grasa con diferentes características. Este hecho fue confirmado por Panford y De Man (1990) al obtener espectros NIR de varios aceites de semillas (palma, girasol, soja, algodón, lino y colza).

Dichos autores indican que la similitud entre espectros no ha de ser considerada como un inconveniente en el desarrollo de aplicaciones NIRS en grasas y aceites y que las pequeñas diferencias observadas en la posición de los picos no son atribuibles a las diferencias de vibración en los enlaces C–H en los diferentes aceites, sino a la longitud de la cadena del ácido graso, al tipo de grupo donde está presente el enlace C–H (metil o metileno) y a la posición de los dobles enlaces en la molécula.

Cho e Iwamoto (1989), tras obtener los espectros de diferentes tipos de aceites y de ésteres metílicos de los ácidos oleico y linoleico, concluyen que las bandas detectadas a 1.710 y 1.725 nm podían ser debidas a los ácidos linoleico y oleico respectivamente.

En el mismo año, Sato *et al.* (1989), publicaron los resultados obtenidos a partir de los espectros de muestras de margarina y mantequilla. A pesar de la gran similitud entre los espectros de ambos productos, fue posible detectar diferencias en las principales bandas de absorción (1.164, 1.160, 2.144 y 2.176 nm) tras someter a los espectros a un tratamiento matemático de segunda derivada. El estudio concluye que las

bandas a 1.164, 2.144 y 2.176 nm podían ser debidas a la absorción de radiación de enlaces C–H asociados a un doble enlace C=C *cis*, mientras que la banda presente a 2.124 nm debía corresponder a la vibración de grupos carbonilo (C=O) presentes en los ésteres de glicerina.

Sinnaeve *et al.* (1990) y Hermida *et al.* (2001) también obtuvieron similares bandas de absorción en los espectros de mantequilla, a pesar de que dichos espectros se obtuvieron en instrumentos de diferente diseño óptico.

Sato *et al.* (1991) completaron sus trabajos iniciales con el estudio y comparación de los espectros NIR de varios aceites y grasas (mantequilla, grasa de leche de cerda, aceite de soja y aceite de palma) con los espectros de grasas elaboradas a partir de mezclas de triglicéridos, en similar proporción a las de los aceites y grasas considerados.

Entre las conclusiones obtenidas por estos autores destacan que las bandas existentes a 1800 nm aportan información sobre ácidos grasos saturados y que cuanto mayor sea el grado de saturación de los ácidos grasos presentes en la muestra, mayor es el valor de absorbancia ($\log 1/R$) a 1.725 nm. Además se observó una excelente similitud entre los espectros de grasa y aquellos generados a partir de mezclas de triglicéridos. Por todo ello, estos autores recomendaron el intervalo de medida entre 1.600 y 2.200 nm como el más adecuado para poder determinar el perfil de ácidos grasos de grasas o aceites.

Gran parte de la asignación de bandas en la región NIR del espectro se ha realizado a partir de cálculos teóricos entre la frecuencia de vibración de bandas de absorción fundamentales en la región MIR y sus correspondientes sobretonos en la región NIR.

Este es el caso de los trabajos de Devaux *et al.* (1993) y Robert *et al.* (1995) en donde se obtuvo la relación entre espectros NIR y MIR de aceites comerciales mediante análisis de correlación canónico. Los resultados mostraron la semejanza entre las representaciones de las dos primeras variables canónicas obtenidas a partir de los espectros MIR y NIR de las mismas muestras, llegando incluso a describir la variabilidad en dichas representaciones de acuerdo con el contenido de C18:1 y C18:2 presente en los diferentes aceites.

Sin embargo, estos autores concluyen que la realización de este análisis de correlación canónico es insuficiente, por si mismo, para obtener una interpretación exacta de los picos en la región NIR, siendo necesario una estimación teórica a partir de las bandas en la región MIR y una posterior comparación con los valores obtenidos por otros autores para realizar una correcta asignación de bandas.

Cobo (1999) y Cobo *et al.* (2001) estudiaron los espectros NIR de aceites de oliva virgen mientras que Indahl *et al.* (1999) interpretaron los espectros de mayonesas y aceites vegetales empleados en la elaboración de estas mayonesas. Todos estos autores llegaron a la conclusión de que los espectros presentaron los picos de absorción en la región NIR a los mismos valores de longitud de onda ya reseñados previamente para grasas y aceites (1.208, 1.724, 1.760, 2.144, 2.308 y 2.348 nm).

Cobo (1999) y Cobo *et al.* (2001), tras estudiar la desviación típica de los valores de absorbancia a cada longitud de onda de aceites de oliva virgen no filtrados, también observan diferentes picos de máxima variación en la región de 1.908 y 2.182 nm que se explicarían por la presencia en dicho aceite de restos de agua y de proteína respectivamente.

Además, estos autores destacan la importancia de la región comprendida entre 400 y 1.100 nm para determinar parámetros de gran interés en aceite de oliva asociados al color y a estados de oxidación. Chen y Chen (1995) ya mostraron la alta correlación existente entre los valores de absorbancia a 850, 870 y 1.310 nm y el color amarillo y entre los valores a 400, 521 y 860 y el color rojo de diferentes aceites vegetales.

Hourant *et al.* (2000) realizaron un extenso trabajo de interpretación de espectros NIR de diferentes tipos de grasas (manteca y grasa de pescado) y aceites (maíz, girasol, avellana, oliva virgen, nuez de Brasil, coco, cacahuete, palma y soja). Las principales conclusiones del trabajo de Hourant *et al.* (2000) fueron que los aceites con alto contenido en ácidos grasos mono y poliinsaturados presentan mayores valores de absorbancia a 1.164 nm que aquellos aceites ricos en ácidos grasos saturados junto con la observación de una alta correlación entre el índice de yodo de los aceites y los valores de absorbancia a 1.720 y 2.148 nm.

Dichos autores también observaron la existencia de una alta correlación entre esta última longitud de onda y el contenido de ácidos grasos poliinsaturados presentes en el aceite, y la presencia de una correlación positiva entre los valores de absorbancia a 1.720 y 1.760 nm y la cantidad de ácidos grasos monoinsaturados analizados.

2.3.2 ANÁLISIS CUANTITATIVO

Una primera observación acerca del análisis cuantitativo realizado mediante espectroscopía NIR en grasas y aceites es el reducido número de publicaciones, comparado con los trabajos realizados en otros productos agroalimentarios tales como cereales, productos cárnicos, forrajes, etc. Sin embargo, a pesar de que el volumen de trabajos es reducido, el número de parámetros analizados es muy amplio.

Garrido *et al.* (en prensa), realizan un listado exhaustivo de analitos y tipos de grasas y aceites para los cuales la bibliografía NIRS recoge información de aplicaciones cuantitativas. Dichos autores recogen más de 50 parámetros analíticos de interés para el control de calidad de diferentes tipos de grasa y aceites a emplear en nutrición humana y animal (Tabla 2.2). Teniendo en cuenta los objetivos de la presente Tesis Doctoral, únicamente se describirán con más detalle aquellos trabajos donde se haya hecho mención a la determinación mediante la tecnología NIRS del contenido en ácidos grasos en diferentes tipos de aceites y grasas.

Así, el primer trabajo recogido que hace referencia a las posibilidades de determinar el contenido en ácidos grasos de un aceite mediante NIRS fue el desarrollado por El-Rafey *et al.* (1988). En este trabajo se analizaron, en un instrumento Neotec 6450 empleando el intervalo de 1.100 a 2.500 nm, 2 colectivos de muestras de aceite de girasol sometidos a diferentes temperaturas de calentamiento (170 y 200°C aprox.). Haciendo uso de una regresión lineal múltiple (RLM) y de un pretratamiento matemático de los espectros mediante una segunda derivada, se desarrollaron ecuaciones de calibración NIRS para la determinación del contenido en C18:2 en las muestras de aceite de girasol.

Tabla 2.2. Aplicaciones NIR en diferentes tipos de grasas y aceites (adaptada de Garrido *et al.*, en prensa).

TIPO DE MUESTRA	DETERMINACIÓN	REFERENCIA
ACEITE DE SOJA	Índice de acidez	Cho <i>et al.</i> (1998)
	Índice de peróxidos	Cho <i>et al.</i> (1998), Yilzid <i>et al.</i> (2001)
	Índice de yodo	Li <i>et al.</i> (1999)
	Ácidos grasos <i>cis</i>	Li <i>et al.</i> (1999)
	Ácidos grasos <i>trans</i>	Li <i>et al.</i> (1999)
	Índice de saponificación	Li <i>et al.</i> (1999)
	Índice de anisidina	Yilzid <i>et al.</i> (2001)
	Índice de dienos	Yilzid <i>et al.</i> (2001)
ACEITE OLIVA VIRGEN	Conjugados	
	Ácidos grasos libres	Garrido <i>et al.</i> (2000)
	K ₂₇₀	Garrido <i>et al.</i> (2000)
	K ₂₃₅	Garrido <i>et al.</i> (2000)
	K ₂₂₅	Garrido <i>et al.</i> (2000)
	Índice de peróxidos	Garrido <i>et al.</i> (2000)
	Polifenoles totales	Garrido <i>et al.</i> (2000)
	Estabilidad oxidativa	Garrido <i>et al.</i> (2000)
	Humedad	Garrido <i>et al.</i> (2000)
	Puntuación de cata	Garrido <i>et al.</i> (2000)
	Pureza	García-Olmo <i>et al.</i> (2001), Wesley <i>et al.</i> (1995, 1996)
Autenticación y/o Clasificación	Bertran <i>et al.</i> (2000), Downey y Flynn (2002), García-Olmo <i>et al.</i> (2001), Wesley <i>et al.</i> (1995, 1996)	
ACEITE DE GIRASOL	Ácido linoleico	El-Rafey <i>et al.</i> (1988), Sato <i>et al.</i> (1995), Pérez-Vich <i>et al.</i> (1998)
	Ácido palmítico	Pérez-Vich <i>et al.</i> (1998)
	Ácido palmitoleico	Pérez-Vich <i>et al.</i> (1998)
	Ácido esteárico	Pérez-Vich <i>et al.</i> (1998)
	Ácido oleico	Pérez-Vich <i>et al.</i> (1998)
	Compuestos polares	El-Rafey <i>et al.</i> (1988)
Compuestos poliméricos	El-Rafey <i>et al.</i> (1988)	
DIVERSOS ACEITES	Índice de acidez	Chen y Chen (1995)
	Índice de peróxidos	Hong <i>et al.</i> (1994), Chen y Chen (1995), Kawano e Iwamoto (1995), Engelsen (1997)
	Índice de yodo	Chen y Chen (1995), Kawano e Iwamoto (1995), Engelsen (1997)
	Fosfolípidos	Chen y Chen (1995)
	Color amarillo	Chen y Chen (1995)
	Color rojo	Chen y Chen (1995)
	Viscosidad	Chen y Chen (1995), Engelsen (1997)
	Ácido oleico	Chen y Chen (1995)
	Ácidos linoleico	Chen y Chen (1995)
	Días de fritura	Engelsen (1997)
	Triglicéridos diméricos y poliméricos	Engelsen (1997)
	Monoglicéridos	Engelsen (1997)
	Diglicéridos	Engelsen (1997)
	Triglicéridos	Engelsen (1997)
	Ácidos grasos libres	Engelsen (1997)
	Índice de anisidina	Engelsen (1997)
	Test oxifrit	Engelsen (1997)
	Vitamina E	Engelsen (1997)
Pureza	Chen y Chen (1995)	
Autenticación y/o Clasificación	Sato (1994), Bewing <i>et al.</i> (1994,1995), Hourant <i>et al.</i> (2000)	
ACEITE DE PALMA	Ácidos grasos libres	Che Man y Moh (1998)

Continuación Tabla 2.2. Aplicaciones NIR en diferentes tipos de grasas y aceites (adaptada de Garrido *et al.*, 2002).

TIPO DE MUESTRA	DETERMINACIÓN	REFERENCIA
ACEITE DE SÉSAMO	Índice de acidez	Ha <i>et al.</i> (1998)
	Índice de peróxidos	Ha <i>et al.</i> (1998)
	Ácido palmítico	Ha <i>et al.</i> (1998)
	Ácido esteárico	Ha <i>et al.</i> (1998)
	Ácido oleico	Ha <i>et al.</i> (1998)
	Ácido linoleico	Ha <i>et al.</i> (1998)
	Pureza	Cho e Iwamoto (1989)
ACEITE DE COLZA	Índice de peróxidos	Dong <i>et al.</i> (1997)
ACEITE DE PESCADO	Ácidos grasos libres	Zhang y Lee (1997)
GRASA DE CERDO	Ácido palmítico	De Pedro <i>et al.</i> (1992), González (1997), Ripoche y Guillard (2001), Pérez-Marín <i>et al.</i> (2002)
	Ácido esteárico	De Pedro <i>et al.</i> (1992), González (1997), Ripoche y Guillard (2001), Pérez-Marín <i>et al.</i> (2002)
	Ácido oleico	De Pedro <i>et al.</i> (1992), González (1997), Ripoche y Guillard (2001), Pérez-Marín <i>et al.</i> (2002)
	Ácido linoleico	De Pedro <i>et al.</i> (1992), González (1997), Ripoche y Guillard (2001), Pérez-Marín <i>et al.</i> (2002)
	Ácido mirístico	González (1997), Pérez-Marín <i>et al.</i> (2002)
	Ácido palmitoleico	González (1997), Pérez-Marín <i>et al.</i> (2002)
	Ácido linolénico	González (1997)
	Ácido araquidónico	González (1997)
	Índice de yodo	Czarwick-Matuszewicz y Korniewicz (1995)
Autenticación y/o Clasificación	Hervás <i>et al.</i> (1994), De Pedro <i>et al.</i> (1995, 2001), González (1997).	
MANTEQUILLA	Humedad	Weaver (1984), Sato <i>et al.</i> (1988), Sinnaeve <i>et al.</i> (1990), Hermida <i>et al.</i> (2001)
	Grasa	Sato <i>et al.</i> (1988), Sinnaeve <i>et al.</i> (1990), Hermida <i>et al.</i> (2001)
	Sólidos no grasos	Sato <i>et al.</i> (1988), Sinnaeve <i>et al.</i> (1990), Hermida <i>et al.</i> (2001)
	NaCl	Sato <i>et al.</i> (1988)
	Pureza	Sato <i>et al.</i> (1989, 1990)
MAYONESA	Autenticación y/o Clasificación	Indahl <i>et al.</i> (1999)
GRASA ANIMAL	Humedad	Valdés y Leeson (1994), Garrido <i>et al.</i> (2002)
	Ácidos grasos libres	Garrido <i>et al.</i> (2002)
	Impurezas	Garrido <i>et al.</i> (2002)
	Insaponificables	Garrido <i>et al.</i> (2002)
	Ácido laurico	Garrido <i>et al.</i> (2002)
	Ácido mirístico	Garrido <i>et al.</i> (2002)
	Ácido palmítico	Garrido <i>et al.</i> (2002)
	Ácido palmitoleico	Garrido <i>et al.</i> (2002)
	Ácido esteárico	Garrido <i>et al.</i> (2002)
	Ácido oleico	Garrido <i>et al.</i> (2002)
	Ácido linoleico	Garrido <i>et al.</i> (2002)
	Ácido margárico	Garrido <i>et al.</i> (2002)
	Ácido margaroleico	Garrido <i>et al.</i> (2002)
	Ácido araquidónico	Garrido <i>et al.</i> (2002)
	Ácido behénico	Garrido <i>et al.</i> (2002)
Energía metabolizable	Valdés y Leeson (1994)	
Energía bruta	Valdés y Leeson (1994)	

Los resultados obtenidos mostraron una alta correlación (coeficientes de correlación superiores a 0,99) entre el contenido en C18:2 obtenido mediante el método de referencia y los valores predichos por NIRS, oscilando los errores típicos de calibración (ETC) entre 0,44 y 0,52% para las diferentes ecuaciones generadas. Sin embargo, en este trabajo no se realiza una validación de dichas ecuaciones, hecho que destacan los propios autores al concluir la necesidad de evaluar las calibraciones obtenidas con otros tipos de aceites sometidos a diferentes temperaturas de calentamiento.

En este mismo tipo de aceite, Sato *et al.* (1995) mostraron la existencia de una alta correlación entre el contenido en C18:2 y los valores espectrales de aceite extraído de semillas de girasol de diferentes variedades. Estos espectros se obtuvieron en un instrumento IA500 (Bran+Luebbe) con un intervalo de medida de 1100 a 2500 nm, empleando la modalidad de transflectancia.

En el colectivo de calibración empleado (n=30), los valores del contenido en C18:2 de las muestras de aceite oscilaron entre 30,6 y 77,0%. Haciendo uso únicamente de la región espectral NIR comprendida entre 1.696 y 1.724 nm, se obtuvo un valor de *R* superior a 0,98 entre los valores de absorbancia en dicha región espectral y el contenido en C18:2, no aportándose en dicho trabajo ninguna estimación del error de predicción de las ecuaciones generadas.

Pérez Vich *et al.* (1998) realizaron un trabajo similar al desarrollado por Sato *et al.* (1995) si bien con diferencias respecto al número de muestras del colectivo de calibración, forma de presentación de muestra para el análisis NIRS y tipo de instrumento empleado. Así, en este trabajo se emplearon 120 muestras de aceite de girasol obtenido a partir de 387 semillas de diferentes variedades. Estas muestras se analizaron en un instrumento Foss NIRSystems 6500 mediante la modalidad de reflectancia, depositando varias gotas de aceite sobre un disco de fibra de vidrio desechable.

Es de destacar en este trabajo el amplio número de muestras de aceite empleadas como colectivo de calibración, las cuales se obtuvieron de semillas de girasol provenientes de un programa de mejora genética. Ello explica la gran variabilidad del contenido en ácidos grasos existente en dicho colectivo. Así, el contenido de C16:0

osciló entre 3,9 y 35,8%, el de C16:1 entre 0,0 y 8,6%, el de C18:0 entre 1,7 y 28,5%, el de C18:1 entre 9,1 y 90,5% y el de C18:2 entre 1,9 y 64,4%.

Dichos autores obtuvieron ecuaciones de calibración NIRS, utilizando como método de regresión el de mínimos cuadrados parciales (RMCPM), no sólo para la determinación del contenido en C18:2 sino también para otros ácidos grasos mayoritarios en este tipo de aceite como son C16:0, C16:1, C18:0 y C18:1. Los valores de R^2 obtenidos en todas las ecuaciones NIRS fueron superiores a 0,97 mientras que los valores de error típico de validación cruzada (*ETVC*) fueron de 1,26% para C16:0, 0,35% para C16:1, 1,22% para C18:0, 1,09% para C18:1 y 0,78% para C18:2. Los resultados obtenidos en este trabajo mostraron la posibilidad de determinar de forma precisa y rápida, mediante la tecnología NIRS, el contenido en ácidos grasos mayoritarios de aceite de girasol.

Chen y Chen (1995) utilizaron 42 muestras de aceite de soja mezclado con aceites de cacahuete, sésamo y girasol y emplearon la modalidad de análisis de transmisión en un instrumento Foss NIRSystems 6500, para la obtención de ecuaciones NIRS de determinación del contenido en C18:1 y C18:2 en dichas mezclas de aceites. La mitad del colectivo inicial fue empleada para el desarrollo de las ecuaciones de calibración y la otra mitad para la validación de dichas ecuaciones. En el desarrollo de las ecuaciones de calibración NIRS, estos autores emplearon regresión lineal múltiple como método de regresión (RLM) a partir de, únicamente, los valores de absorbancia a 3 longitudes de onda (1.720, 1.750 y 2.150 nm).

Los valores de R^2 y de error típico de predicción (*ETP*) para el contenido en C18:1 fueron de 0,86 y 1,21% respectivamente mientras que para el porcentaje de C18:2 fueron de 0,83 y 1,36% respectivamente. Los autores concluyeron que los valores del estadístico *RPD* (cociente entre la desviación típica del colectivo de validación y el *ETP*) fueron 7,2 para el ácido C18:1 y 6,8 para el ácido C18:2, siendo dichos valores muy superiores a los valores recomendados para este estadístico por Williams y Sobering (1993).

Ha *et al.* (1998) estudiaron el potencial de la técnica NIRS para la determinación de varios parámetros de calidad en el aceite de sésamo, y en particular para la determinación del contenido en los ácidos grasos C16:0, C18:0, C18:1 y C18:2.

Para ello, a partir de 113 muestras de aceite de sésamo obtenidas de diferentes regiones de Corea, se obtuvieron los espectros NIR en dos equipos diferentes: un instrumento de filtros (IA450) y un instrumento con monocromador (IA500).

En general, mediante la utilización del equipo monocromador (IA500) se obtuvieron ecuaciones de calibración con valores R^2 más altos y ETC más bajos que las desarrolladas a partir del instrumento de filtros (IA450). Sin embargo, tras realizar una validación externa, los resultados no fueron tan concluyentes. Así, los valores de ETP obtenidos con las ecuaciones desarrolladas en el instrumento IA450 e IA500 fueron, respectivamente, 0,31 y 0,33% para el ácido C16:0, 0,32 y 0,14% para C18:0, 0,68 y 0,52% para el ácido C18:1 y 0,99 y 1,04% para C18:2. Los valores de ETP obtenidos para los ácidos grasos C18:1 y C18:2 fueron inferiores a los obtenidos por Chen y Chen (1995), si bien no es posible hacer una comparación en términos relativos al no indicar el trabajo de Ha *et al.* (1998) los estadísticos (media, desviación típica o rango) de los colectivos de calibración y validación.

Después de comparar los resultados obtenidos en ambos instrumentos NIR, Ha *et al.* (1998) recomiendan, a modo de conclusión, el uso del instrumento equipado con monocromador (IA500) para determinar el contenido en ácidos grasos en aceite de sésamo.

En los trabajos expuestos previamente, se han evaluado las posibilidades de uso de la tecnología NIRS para la determinación de ácidos grasos en aceites de semillas. Sin embargo, resulta de gran interés para esta tesis la revisión de aquellos trabajos en los que se hayan desarrollado asimismo ecuaciones para la predicción de ácidos grasos en grasa animal y, en particular, en grasa de cerdo.

De Pedro *et al.* (1992) utilizaron un colectivo de 118 muestras de grasa fundida de cerdo Ibérico, analizadas mediante la modalidad de transmisión en un instrumento monocromador PSCO 6250 y empleando para ello una cápsula diseñada por dichos autores con un paso óptico de 1 mm. En este trabajo, pionero en la determinación de varios ácidos grasos de forma conjunta en una grasa o aceite, se desarrollaron ecuaciones NIRS para la determinación del contenido en los ácidos grasos mayoritarios (C16:0, C18:0, C18:1 y C18:2) en grasa de cerdo Ibérico, en el intervalo de medida espectral de 1.100 a 2.500 nm y empleando tanto la regresión lineal múltiple

(RLM) como regresión por mínimos cuadrados parciales (RMCP). El conjunto inicial de muestras de grasa líquida fue dividido en un colectivo de calibración ($n=83$) y otro para la validación de las ecuaciones desarrolladas ($n=36$).

Las ecuaciones obtenidas mediante regresión RMCP mostraron mayor exactitud y precisión que las desarrolladas mediante RLM. Así, los valores de *ETC* para las ecuaciones desarrolladas mediante regresión RLM o RMCP fueron, respectivamente, 0,65 y 0,40% para el ácido C16:0, 0,66 y 0,44% para C18:0, 0,80 y 0,59% para el ácido C18:1 y 0,42 y 0,35% para C18:2. Los valores del coeficiente de correlación oscilaron entre 0,83 y 0,95 para las ecuaciones desarrolladas mediante regresión RLM y entre 0,93 y 0,98 para las ecuaciones RMCP. Indicar que, para las ecuaciones generadas mediante RLM, las longitudes de onda seleccionadas para todos los ácidos grasos estaban comprendidas, fundamentalmente, en las regiones de 2.120 a 2.190 nm y entre 2.440 y 2.490 nm. En el caso del ácido C18:1, también se seleccionaron longitudes de onda en la región de 1700 nm.

Estas diferencias entre los estadísticos de las ecuaciones generadas mediante diferentes algoritmos de regresión también se mantuvieron al validar las ecuaciones con un colectivo de muestras no empleado durante el desarrollo de las calibraciones. Los valores del error típico de predicción (*ETP*) fueron ligeramente superiores a los valores de *ETC*, siendo menores las diferencias entre los valores de *ETC* y *ETP* en las ecuaciones generadas mediante el algoritmo RMCP. Así, los valores de *ETP* para las ecuaciones desarrolladas mediante regresión RLM o RMCP fueron, respectivamente, 0,69 y 0,43% para el ácido C16:0, 0,66 y 0,58% para C18:0, 1,05 y 0,59% para el ácido C18:1 y 0,39 y 0,36% para C18:2.

Por todo ello, los autores de este trabajo concluyen que la tecnología NIRS puede ser empleada para predecir el contenido en ácidos grasos mayoritarios en grasa fundida de cerdo Ibérico obtener resultados con elevada exactitud, precisión y rapidez. Sin embargo, también indican que estos resultados podrían mejorarse siempre y cuando se minimizaran los errores en la preparación de la muestra para su análisis.

Posteriormente, González (1997) partiendo de espectros NIR obtenidos tanto en grasa fundida como en grasa picada de 70 muestras analizadas mediante reflectancia en un instrumento IA500 (Bran+Luebbe), con un rango de medida de 1.100

a 2.500 nm, desarrolló ecuaciones NIR para la determinación, no sólo de los ácidos grasos mayoritarios (C16:0, C18:0, C18:1 y C18:2) si no también de alguno de los minoritarios (C14:0, C16:1, C18:3 y C20:0) en grasa subcutánea y perirrenal de cerdo Ibérico. También se consideraron otras variables como fueron el porcentaje total de ácidos grasos saturados, insaturados, monoinsaturados, poliinsaturados y las relaciones entre ellos (% insaturados / % saturados; % monoinsaturados / % saturados; % poliinsaturados / % saturados; % monoinsaturados / % poliinsaturados).

Para la generación de las ecuaciones de calibración se empleó el algoritmo de regresión de mínimos cuadrados parciales (RMCP), considerando la influencia del pretratamiento de corrección de los fenómenos de *scatter* o de radiación dispersa (*MSC*) sobre la precisión y exactitud de las ecuaciones. No se realizó ningún tratamiento de validación, ya sea cruzada o externa, empleando todo el colectivo inicial (n=70) como colectivo de calibración.

En general, las ecuaciones de calibración que utilizaban la forma de presentación de grasa fundida y corrección *MSC*, fueron las que presentaron mayor exactitud y precisión. Así, los valores de R^2 y *ETC* en las mejores ecuaciones desarrolladas para los ácidos grasos mayoritarios en grasa de cerdo Ibérico fueron, respectivamente, 0,91 y 0,83% para el ácido C16:0, 0,92 y 0,69% para C18:0, 0,93 y 1,11% para el ácido C18:1 y 0,91 y 0,51% para C18:2. En el caso de las ecuaciones generadas para los ácidos grasos minoritarios, estas presentaron inferior exactitud que las desarrolladas para los ácidos grasos mayoritarios. Así, los valores de R^2 y *ETC* fueron, respectivamente, de 0,40 y 0,17% para el ácido C14:0, 0,41 y 0,40% para C16:1, 0,24 y 0,19% para el ácido C18:3, 0,10 y 0,36% para C20:0.

En el desarrollo de ecuaciones NIRS para la determinación del contenido en la suma de varios ácidos grasos (% saturados, % insaturados, % monoinsaturados, % poliinsaturados) y su relación entre ellos, se obtuvo una excelente correlación entre los valores de referencia y los datos espectrales NIR, con valores de R^2 superiores a 0,90 a excepción de la ecuaciones generada para determinar la relación monoinsaturados/poliinsaturados, donde el valor de R^2 fue ligeramente superior a 0,75.

Las ecuaciones generadas para grasa perirrenal, tanto picada como fundida y con tratamiento *MSC* de corrección de la radiación dispersa presentaron menor

exactitud y precisión que las generadas para grasa subcutánea, con valores de R^2 inferiores a 0,90 para todas las variables consideradas y valores de ETC , en general, un 20% superiores a las anteriormente mencionadas.

A modo de conclusión, González (1997) indicó que la tecnología NIRS podía aplicarse con éxito para el análisis de composición de ácidos grasos mayoritarios y minoritarios en grasa subcutánea de cerdo Ibérico, recomendándose el uso de un tratamiento MSC sobre los espectros NIR originales para la obtención de los modelos de regresión de variables cuantitativas.

Recientemente, Pérez-Marín *et al.* (2002), aplicaron la metodología puesta a punto para el análisis NIRS de grasa fundida de cerdo Ibérico, como parte del presente trabajo de Tesis Doctoral a un colectivo de calibración de 188 muestras de grasa subcutánea de cerdo Ibérico, analizadas en modo fundida mediante la modalidad de doble transmisión con una cápsula de fondo de oro reflectante. El instrumento empleado fue un Foss NIRSystems 6500 SYI equipado con módulo de giro, considerando un intervalo de medida de 1.100 a 2.500 nm. Una vez obtenidas las ecuaciones NIRS de determinación del contenido en ácidos grasos mayoritarios (C16:0, C18:0, C18:1 y C18:2), se evaluó su capacidad predictiva sobre un colectivo de validación compuesto por 12 muestras que cubrían la mayor parte del intervalo de variabilidad existente en la composición de grasa de cerdo Ibérico.

Los valores de ETP obtenidos en dicho trabajo fueron excelentes, siendo del mismo orden de magnitud que los valores del error del método de referencia (cromatografía de gases). Así, dichos valores fueron de 0,28% para el ácido C16:0, 0,27% para C18:0, 0,20% para C18:1 y 0,16% para C18:2. Además, los valores de R^2 resultantes de la validación externa fueron de 0,94 para el ácido C18:2, 0,97 para C16:0 y 0,99 para C18:0 y C18:1.

Garrido *et al.* (2002) desarrollaron ecuaciones para la determinación de varios parámetros de interés en grasas destinadas a alimentación animal, utilizando la misma modalidad de análisis, el mismo tipo de cápsula (cápsula de fondo reflectante de oro) y el mismo instrumento (Foss NIRSystems 6500 SYI equipado con módulo de giro) que los utilizados por Pérez-Marín *et al.* (2002).

Entre los parámetros estudiados se ha de destacar el desarrollo de ecuaciones NIRS tanto para la predicción del porcentaje en cada uno de los ácidos grasos mayoritarios (C16:0, C18:0, C18:1 y C18:2) como de los minoritarios (C12:0, C14:0, C16:1, C17:0, C17:1, C20:0 y C22:0). Como colectivo de calibración se emplearon 61 muestras de grasa de diferentes características, estimando la capacidad predictiva de las ecuaciones mediante un proceso de validación cruzada. Para el desarrollo de dichas ecuaciones, se consideró la región de 1.100 a 2.500 nm.

A pesar de que estas muestras pueden ser, a priori y en base a su aspecto, muy diferentes a una grasa de cerdo Ibérico, ambos tipos de muestras pueden llegar a tener una composición en ácidos grasos mayoritarios similar. Así, las muestras empleadas en dicho trabajo presentaron unos valores medios de 21,7% en C16:0, 13,7% en C18:0, 37,7% en C18:1 y 13,6% en C18:2. Sin embargo, las diferencias tanto en medias como en intervalos de variación pueden ser mayores en algunos ácidos grasos minoritarios como C14:0 (media = 1,95%), C17:0 (media = 0,67%) o C17:1 (media = 0,42%).

Los valores de *ETVC* de las ecuaciones generadas para los ácidos grasos mayoritarios fueron de 0,76% para C16:0, 0,74% para C18:0, 0,97% para C18:1 y 0,65% para C18:2, con valores de r^2 en todos los casos superiores a 0,92. Para los ácidos grasos minoritarios, los valores de *ETVC* fueron de 0,01% para C12:0, 0,13% para C14:0, 0,33% para C16:1, 0,06% para C17:0, 0,04% para C17:1, 0,02% para C20:0 y 0,04% para C22:0. En este caso, los valores de r^2 fueron inferiores a los obtenidos para los ácidos grasos mayoritarios, si bien dicho estadístico fue superior a 0,85 en todos los ácidos grasos minoritarios a excepción de C12:0 y C17:1 donde presentaron valores de 0,67 y 0,77 respectivamente.

Todos los trabajos expuestos previamente han sido realizados sobre muestras de aceite o grasas animales en estado líquido. Recientemente Ripoche y Guillard (2001) han publicado un trabajo en el que se ha evaluado el desarrollo de ecuaciones NIR para la determinación de ácidos grasos individuales tanto en muestras de tejido adiposo subcutáneo de cerdo como en la grasa líquida extraída de dichas muestras. En el primer caso (tejido adiposo), la modalidad de análisis empleada fue de reflectancia difusa mientras que en el segundo (grasa líquida) se hizo uso de la

modalidad de transmisión. Los espectros se obtuvieron mediante un espectrofotómetro con transformadas de Fourier Bomem MB100 en el intervalo de 900 a 2.500 nm.

A partir de un colectivo de 50 muestras de tejido adiposo y su correspondiente grasa fundida, Ripoché y Guillard (2001) generaron ecuaciones, mediante regresión RMCP, para determinar el porcentaje en ácidos grasos saturados, monoinsaturados, poliinsaturados y los ácidos grasos individuales C16:0, C18:0, C18:1 y C18:2. La precisión y exactitud de las ecuaciones generadas se evaluó mediante la validación externa de un colectivo de 40 muestras de tejido adiposo y de grasa fundida.

Los valores de *ETP* obtenidos en las ecuaciones generadas para grasa fundida fueron de 0,49% para C16:0, 0,63% para C18:0, 0,58% para C18:1 y 0,51% para C18:2, oscilando los valores de r^2 entre 0,81 y 0,98. Las ecuaciones desarrolladas en tejido adiposo presentaron unos errores muy superiores, incluso inaceptables, para los ácidos grasos C16:0 y C18:0. Los valores de *ETP* y r^2 fueron 4,3% y 0,72 para C18:1 y 2,7% y 0,79 para C18:2.

2.3.3 ANÁLISIS CUALITATIVO

Han sido varios los estudios dirigidos a evaluar la potencialidad de la tecnología NIRS para resolver problemas relacionados con clasificación, autenticación y adulteración en aceites y grasas (Tabla 2.2). A continuación se sintetizarán estos trabajos agrupando su discusión en función del tipo de producto sobre el que se ha trabajado (aceites vegetales y grasas animales).

2.3.3.1 *Aceites vegetales*

Uno de los primeros trabajos de clasificación de aceites vegetales a partir de la información espectral *per se* fue el realizado por Sato (1994). Este autor analizó un total de 32 muestras de 9 tipos de aceites comerciales diferentes (soja, maíz, algodón, oliva, arroz, colza, sésamo, coco y cacahuete) mediante la modalidad de doble transmisión en un monocromador IA 500 (Bran+Luebbe). El paso óptico de la cápsula utilizada fue de 0,3 mm. En este trabajo se consideraron tanto los datos espectrales brutos ($\log 1/R$) como derivados (2ª derivada) para el intervalo de 1.600 a 2.200 nm.

Indicar que todo el colectivo fue empleado en la construcción del modelo de clasificación, no destinando ninguna muestra para su validación.

Sobre estos datos espectrales NIR, se aplicó un análisis de componentes principales, proyectando a continuación las muestras en planos definidos por las primeras componentes principales. Para dicho cálculo de componentes principales, los autores recomiendan el uso de la matriz de covarianzas, frente a la matriz de correlación al minimizar la primera matriz la influencia que puedan tener sobre el modelo aquellas regiones espectrales con valores de absorbancia próximos a cero.

A partir de la observación de dichas proyecciones y teniendo en cuenta la composición en ácidos grasos de los aceites analizados, se observó que los 9 grupos de aceites vegetales resultaban ordenados a lo largo del eje definido por la 2ª componente principal. Así, cuando se representaron los espectros brutos ($\log 1/R$) sobre un plano definido por la 2ª y 3ª componente principal, los aceites con mayor grado de insaturación (soja, maíz, algodón, colza y sésamo) se localizaron en el extremo positivo del eje definido por la 2ª componente principal mientras que los aceites con menor grado de insaturación (arroz, cacahuete, oliva y coco) se proyectaron hacia valores más negativos de dicho eje.

Al representar la segunda derivada de los datos espectrales en un plano definido por la 1ª y 2ª componente principal, los resultados fueron similares, si bien, al emplear datos NIR con derivación (2ª derivada), se logró una mayor separación y clasificación de los diferentes tipos de aceites que al emplear datos espectrales sin derivar ($\log 1/R$).

Utilizando también datos derivados, Sato (1994) llevó a cabo una interpretación química de los vectores propios obtenidos al realizar el análisis de componentes principales, indicando que las 3 primeras componentes principales explicaban el 99,87% de la variabilidad existente en los datos espectrales del colectivo de aceites. Tras esta interpretación observó que la primera componente principal no aportaba información sobre la composición en ácidos grasos de las muestras de aceites. Sin embargo, la segunda componente principal se correlacionó con el grado de insaturación existente en el aceite mientras que la tercera componente principal aportó información no solo del grado de insaturación existente sino también de la calidad

composicional de los ácidos grasos insaturados (C18:1, C18:2 y C18:3) al correlacionarse con la relación entre ellos.

Bewig *et al.* (1994 y 1995) estudiaron asimismo la potencialidad de la tecnología NIRS para distinguir entre diferentes tipos de aceites vegetales así como para clasificar muestras desconocidas. Emplearon un total de 46 muestras de 4 tipos diferentes de aceites vegetales (cacahuete, soja, algodón y colza) analizadas por triplicado en un instrumento IA 500 (Bran+Luebbe) mediante la modalidad de transflectancia, empleando un paso óptico de 1 mm. Este conjunto de datos espectrales fue dividido en 2: uno de ellos empleado como colectivo de aprendizaje y otro como colectivo de validación. Además, se consideró otro colectivo de validación a partir de muestras de aceite de girasol, maíz y oliva no considerados en el colectivo de aprendizaje.

A partir de los espectros con un tratamiento de 2ª derivada, se obtuvieron modelos discriminantes empleando varias longitudes de onda seleccionadas a partir de la suma del cuadrado de la inversa de la distancia de Mahalanobis entre 2 grupos. El mejor modelo discriminante desarrollado para diferenciar los 4 tipos de aceites empleó 4 longitudes de onda (1.704, 1.802, 1.816 y 2.110 nm).

Este modelo se aplicó sobre el primer colectivo de validación, empleando como criterio de clasificación que la distancia de Mahalanobis de la muestra a clasificar al centro de una clase fuera inferior a 3. En caso de que la distancia de Mahalanobis fuera superior a 3 para todas las clases, la muestra sería asignada a aquel grupo al cual presentara una menor distancia. Con este criterio, únicamente 2 de las 13 muestras del primer colectivo de validación fueron clasificadas erróneamente, llegando los autores a emitir la hipótesis de que dicho error de clasificación fue menor tras estudiar varios parámetros químicos (índice de yodo y composición en ácidos grasos) de los aceites analizados. Al aplicar el modelo discriminante sobre el segundo colectivo de validación considerado, todas las muestras presentaron una distancia de Mahalanobis superior a 3 a todas las categorías.

Hourant *et al.* (2000), además del análisis espectral descrito en el apartado 2.3.1, intentaron demostrar que era posible discriminar diferentes tipos de grasas y aceites mediante el uso de varias longitudes de onda del rango NIR. Para ello,

emplearon un total de 104 muestras de grasas y aceites con diferente grado de insaturación. Un 62% de dicho total fue empleado como colectivo de aprendizaje y el 38% restante como colectivo de validación.

A partir de dicho colectivo de aprendizaje y mediante análisis discriminante lineal diseñó un organigrama de discriminación mediante el cual clasificar las muestras analizadas en las siguientes categorías: grasas saturadas, aceite de maíz, soja, girasol, colza, cacahuete, girasol con alto contenido en ácido oleico y aceite de oliva virgen. Cada una de las funciones de discriminación desarrolladas y que constituían el organigrama de clasificación se generó a partir de entre 3 y 6 longitudes de onda seleccionadas de la región NIR.

Los resultados de clasificación del colectivo empleado para la validación oscilaron en función del tipo de aceite o grasa. Así, el 99% de las muestras de aceites de soja, girasol y oliva fueron correctamente clasificados por los modelos. Sin embargo, estos porcentajes fueron inferiores en el caso de los aceites de maíz (66,7%) o en los de girasol con alto contenido en ácido oleico (50%).

Otros trabajos realizados sobre aceites vegetales han estudiado la tecnología NIRS como herramienta de detección del grado de pureza o del porcentaje de adulteración con otros aceites de semillas. Un ejemplo de ellos son los trabajos realizado sobre aceite de sésamo por Cho e Iwamoto (1989), en aceite de soja por Chen y Chen (1995) y aceite de oliva por Wesley *et al.* (1995, 1996) y Garcia Olmo *et al.* (2001).

Cho e Iwamoto (1989) evaluaron las posibilidades de determinar el porcentaje de aceite de sésamo en mezclas con aceite de maíz, a partir de sus espectros NIR. Para ello, emplearon 22 espectros de mezclas de aceites para el desarrollo de ecuaciones y 12 espectros para su validación, obtenidos todos ellos mediante un instrumento de Bran+Luebbe IA 500 mediante la modalidad de transreflectancia con 0,2 mm de paso óptico. Las ecuaciones desarrolladas mediante regresión RLM presentaron, según los propios autores, una aceptable precisión para determinar la proporción de aceites en las mezclas, obteniéndose unos valores de 0,99 y 1,21% para los estadísticos R y ETP respectivamente en la ecuación generada a partir de un máximo de 4 longitudes de onda (1.700, 1.720, 1.748 y 2.232 nm).

Chen y Chen (1995) emplearon la relación entre el contenido en ácido C18:1 y C18:2 de 42 mezclas de aceite de soja con aceites de cacahuete, sésamo y girasol. Para ello, hicieron uso de las ecuaciones descritas en el apartado 2.3.2, para determinar cada uno de estos ácidos grasos, a partir de los espectros NIR obtenidos mediante la modalidad de transmisión de las mezclas de aceites. A continuación, calcularon la relación entre ambos ácidos grasos y establecieron una regresión entre dicha relación y el porcentaje de pureza del aceite de soja. Los coeficientes de correlación obtenidos fueron de 0,98 para las mezclas con aceite de girasol, 0,94 en el caso de emplear aceite de sésamo y 0,97 en las mezclas con aceite de cacahuete.

Wesley *et al.* (1995) analizó, mediante la modalidad de transmisión en un instrumento Foss NIRSystems 6500, un total de 319 muestras de aceite de oliva virgen y extra virgen adulterado (entre un 0 y 30% en p/p) con aceite de girasol, maíz y aceite de oliva refinado. A partir de un colectivo de 250 espectros de las muestras iniciales se desarrollaron ecuaciones RMCP para la predicción del porcentaje de adulteración, las cuales se aplicaron sobre 60 espectros del colectivo inicial obteniéndose un valor de *ETP* del 1,78%. Al validar mediante estas ecuaciones 40 espectros de otras muestras de aceites no incluidas en el colectivo de calibración, el valor de *ETP* fue de 4,5%.

Con este mismo colectivo de 250 espectros se realizó un análisis de componentes principales para identificar el tipo de agente adulterante empleado, obteniéndose, como mejor resultado, un 75% de las muestras correctamente clasificadas. Por ello, Wesley *et al.* (1995) concluyen que tanto los resultados de predicción del grado de adulteración como el modelo para identificar el tipo de aceite empleado en la adulteración deberían optimizarse empleando un colectivo inicial que tuviera un mayor número de muestras de aceite de oliva en las que se reflejara la amplia variabilidad de producto.

En un estudio posterior, estos mismo autores (Wesley *et al.*, 1996) generaron ecuaciones de predicción del grado de adulteración y modelos de clasificación de diferentes tipos de aceites a partir de 23 mezclas de aceite de oliva con aceite de colza, 21 mezclas con aceite de soja, 23 con aceite de girasol y 6 muestras de aceites de oliva virgen extra, con porcentajes de adulteración de aceites de semillas en aceite de oliva entre el 5% y 95%. Al analizar dichas muestras de aceite por duplicado, mediante la misma modalidad en Wesley *et al.* (1995), se constituyó un colectivo inicial

de 152 espectros, de los cuales, 20 fueron seleccionados aleatoriamente para validar las ecuaciones y los modelos obtenidos.

La validación externa de la ecuación generada, mediante el algoritmo RMCPM, para predecir el porcentaje de aceite de oliva en las mezclas de aceites, presentó unos valores de R^2 y ETP de 0,99 y 2,77% respectivamente. Sin embargo, este valor de ETP disminuyó sensiblemente hasta un valor de 0,68% cuando del colectivo de validación se eliminaron los espectros de aceites de oliva puros, dada su escasa representación en el colectivo empleado durante la calibración.

Los resultados obtenidos en este mismo trabajo al desarrollar modelos de clasificación fueron también muy satisfactorios. Así, empleando como variables las componentes principales de los datos espectrales NIR, se generaron modelos discriminantes para identificar el tipos de agente adulterante empleado. Estos modelos clasificaron correctamente el 90% de las muestras del colectivo de validación.

Un trabajo similar al descrito con anterioridad fue el desarrollado por García Olmo *et al.* (2001) si bien, se evaluó el potencial de la técnica NIRS para detectar adulteraciones en aceite de oliva a menores concentraciones (hasta 0,5%) y otros agentes adulterantes como el aceite de girasol con alto contenido en ácido oleico. Dichos autores desarrollaron mezclas binarias de aceite de oliva virgen extra con aceite de soja, colza, girasol y girasol con alto contenido en ácido oleico, en porcentajes entre 0,5% hasta 100% de agente adulterante, obteniendo 82 espectros tras analizar dichas muestras por duplicado mediante la modalidad de doble transmisión.

Las ecuaciones, generadas mediante regresión RMCPM, para predecir el porcentaje de aceite adulterante en las mezclas binarias, presentaron valores de R^2 de 0,99 en todos los casos. Los valores de $ETVC$ obtenidos fueron de 0,39% para la predicción del porcentaje de aceite de soja, 0,21% para la ecuación de aceite de colza, 0,32% para la de aceite de girasol y 0,85% para la predicción del porcentaje de aceite de girasol con alto contenido de ácido oleico.

Los modelos de clasificación desarrollados por García Olmo *et al.* (2001) para detectar el tipo de aceite adulterante, se generaron mediante análisis discriminante cuadrático, empleando como variables las componentes principales de los datos espectrales NIR. Dado el pequeño número de muestras disponibles, los autores

emplearon la validación cruzada completa como técnica de validación de los modelos desarrollados, obteniéndose un error del 7,1%. La mayor parte de este error fue debida a la incorrecta clasificación de muestras con bajos porcentajes de adulteración (inferiores al 1%), no detectándose ninguna clasificación de muestras de aceite de oliva adulterado en la categoría de muestras de aceite de oliva virgen extra.

Otros trabajos de autenticación y clasificación de aceite de oliva mediante la técnica NIRS han ido orientados hacia la diferenciación de aceites de diferente origen, como son los desarrollados por Bertran *et al.* (2000) y Downey y Flynn (2002).

Así, Bertran *et al.* (2000) estudiaron la posibilidad de discriminar aceites de oliva pertenecientes a dos Denominaciones de Origen de España como son “Siurana” y “Les Garrigues”, cercanas en cuanto a su localización geográfica pero con diferentes sistemas de producción oleícola. Para ello se emplearon un total de 85 muestras de ambas denominaciones de origen, 29 de las cuales fueron seleccionadas para validar externamente los modelos generados.

Todas las muestras fueron analizadas, mediante la modalidad de transmisión con una cubeta de 4 mm de paso óptico, en un instrumento Foss NIRSystems 6500. Debido al paso óptico empleado, se produjo la saturación de los detectores las longitudes de onda próximas a 2.500 nm, por lo que la región espectral considerada en este estudio cualitativo fue entre 1.100 y 2.260 nm.

Estos autores compararon los modelos de autenticación desarrollados mediante dos métodos de clasificación (análisis mediante redes neuronales y regresión logística), empleando en ambos casos como variables las componentes principales de los datos espectrales NIR. En los modelos desarrollados mediante redes neuronales, el 100% de las muestras de colectivo de validación externa resultaron correctamente clasificadas mientras que, en los modelos generados mediante regresión logística, 2 espectros de muestras de aceites pertenecientes a la Denominación de Origen “Siurana” fueron incorrectamente clasificadas dentro de la categoría “Les Garrigues”.

Por todo ello, los autores concluyen que los resultados de clasificación son satisfactorios, teniendo en cuenta las semejanzas existentes entre los aceites pertenecientes a ambas Denominaciones de Origen. Además, recomiendan el uso de

regresión logística para el desarrollo de los modelos frente al análisis por redes neuronales dada la menor complejidad de los modelos.

Un trabajo similar al anterior fue el estudio preliminar desarrollado por Downey y Flynn (2002) con el objetivo de evaluar las posibilidades de la técnica NIRS para discriminar aceites de oliva de regiones próximas del área mediterránea como son Creta, Peloponeso y otras islas griegas. Para ello emplearon un total de 64 espectros de muestras de aceites, analizadas mediante la modalidad de doble transmisión (0,5 mm de paso óptico) en un instrumento Foss NIRSystems 6500. Emplearon como método de clasificación en las 3 categorías de aceites el análisis discriminante factorial, considerando la mitad de los espectros como colectivo de calibración y la otra mitad para la validación de los modelos generados.

Los mejores resultados se obtuvieron al desarrollar los modelos a partir de los datos espectrales de la región del visibles y NIR, oscilando los errores de clasificación de las muestras de validación entre un 6,1% y 26,3%.

2.3.3.2 Grasas

Sato *et al.* (1989, 1990) realizaron el primer estudio sobre el uso de la tecnología NIRS para detectar adulteraciones de margarina en mantequilla. Para ello se basaron en la información espectral de ambos tipos de productos y, en particular, en las absorbancias obtenidas a 1.164, 1.660, 2.124 y 2.176 nm como descriptoras de la composición en ácidos grasos de un determinado tipo de grasa. Dichos autores consideraron que una adición de margarina a mantequilla provocaría una alteración en los valores espectrales obtenidos a dichas longitudes de onda.

Por ello, realizaron mezclas binarias de margarina en mantequilla a diferentes porcentajes en peso (0% a 100%). Tras obtener los espectros NIR de dichas mezclas, lograron obtener coeficientes de correlación superiores a 0,98 entre el porcentaje de adulteración y la segunda derivada de los valores de absorbancia del intervalo espectral comprendido entre 2.124 y 2.144 nm. Además, los autores de este trabajo concluyen que la espectroscopía NIR puede ser empleada no sólo para detectar adulteraciones en mantequilla sino también para poder cuantificar dicha adulteración hasta porcentajes del orden del 3%.

Más recientemente, Indahl *et al.* (1999) publicaron un trabajo acerca del uso de la información espectral NIR “*per se*” para la diferenciación de mayonesas obtenidas a partir de seis tipos de aceites vegetales (aceite de soja, girasol, colza, oliva, maíz y semilla de uva). Para ello, obtuvieron los espectros NIR por triplicado de 54 muestras de mayonesa, empleando para ello un instrumento Bran+Luebbe IA 500. Con ello se logró un colectivo inicial de 162 espectros.

Los autores de este trabajo evaluaron tres métodos de clasificación diferentes como son el análisis discriminante lineal, análisis discriminante cuadrático y análisis discriminante mediante regresión. Estos algoritmos se aplicaron sobre tres tipos de variables: componentes principales de los datos espectrales, componentes RMCP2 de datos espectrales y una selección de longitudes de onda de los datos espectrales originales. La evaluación de los modelos se realizó mediante validación cruzada.

De entre todos los modelos generados, los mejores resultados se obtuvieron al emplear análisis discriminante lineal sobre los datos de componentes principales o componentes RMCP de los espectros originales. En este caso se clasificaron correctamente 161 espectros de un total de 162 espectros. Al aplicar un análisis discriminante basado en regresión sobre las componentes principales o componentes RMCP2, el número de espectros correctamente clasificados osciló entre 131 y 158 sobre un total de 162.

En el caso de emplear una selección de longitudes de onda como variables para el desarrollo de los modelos de clasificación, también se obtuvieron mejores resultados en los modelos generados mediante análisis discriminante lineal (entre 153 y 160 muestras correctamente clasificadas) que en aquellos obtenidos mediante regresión (entre 126 y 148 muestras correctamente clasificadas).

Indahl *et al.* (1999) profundizaron aun más en el estudio de clasificación de mayonesas. Para ello obtuvieron los espectros NIR de los diferentes tipos de aceites empleados previamente en la fabricación de las mayonesas y realizaron un análisis discriminante a partir de sus componentes principales. La representación de las dos primeras variables canónicas resultantes del análisis discriminantes desarrollado con los espectros de aceites y mayonesas fueron muy similares. Al mismo tiempo, esta representación fue análoga a la obtenida al representar los valores del ácido graso C18:1

vs. C18:2 para los diferentes tipos de aceites. Por todo ello, los autores concluyeron que la clasificación de las muestras de mayonesa se basó en el tipo de aceite empleado y que, al mismo tiempo, dicha clasificación es realizada fundamentalmente en función del contenido en los ácidos grasos C18:1 y C18:3.

Varios autores han evaluado las posibilidades de aplicación de la técnica NIRS como herramienta de clasificación de muestras de grasa de cerdo Ibérico en función del régimen alimenticio empleado durante su etapa de cebo (Hervás *et al.*, 1994; De Pedro *et al.*, 1995 y 2001; González, 1997).

Los primeros trabajos realizados acerca del uso de la información espectral NIR *per se* como herramienta de autenticación de muestras de grasa de cerdo Ibérico con diferente régimen alimenticio fueron los realizados por Hervás *et al.*, 1994 y De Pedro *et al.*, 1995. Dichos autores emplearon un mismo colectivo inicial formado por 118 muestras de grasa líquida de cerdos Ibéricos sometidos a tres regímenes alimenticios con diferente aporte de bellotas en su dieta.

Los espectros NIR de dichas muestras se obtuvieron, mediante la modalidad de transmisión en el rango de 1.100 a 2.500 nm, mediante un instrumento PSCO 6250. De este colectivo inicial se seleccionaron 64 muestras como grupo de aprendizaje mediante el algoritmo SELECT incluido en el programa ISI. El resto de muestras (54) fue empleado como colectivo de validación de los modelos generados.

A partir de los valores de componentes principales de los datos espectrales originales del colectivo de aprendizaje, se generaron modelos de clasificación mediante análisis discriminante lineal y análisis por redes neuronales. Además, dada la correlación existente en los cerdos Ibéricos entre el régimen alimenticio empleado y la composición en ácidos grasos de su grasa, también se generaron modelos de clasificación para las mismas muestras, mediante análisis discriminante lineal y empleando como variables las componentes principales de los valores de composición de 11 ácidos grasos obtenidos mediante cromatografía de gases.

Al emplear como variable únicamente la información espectral NIR, los resultados desarrollados por Hervás *et al.* (1994) y De Pedro *et al.* (1995) mostraron que el error de clasificación obtenido con el modelo generados mediante análisis por redes neuronales (3,7%) fue inferior al calculado mediante análisis discriminante lineal

(11,3%). Es de descartar que, en todos los modelos generados, ninguna de las muestras de animales que no tuvieron aporte de bellota en su dieta fueron clasificadas erróneamente en el resto de categorías.

Al emplear como variables las componentes principales de la composición en ácidos grasos, el modelo discriminante presentó un error de clasificación del 15,7%. Por ello, el error de clasificación de cualquiera de los modelos generados a partir de la información espectral NIR *per se* fue inferior al error obtenido cuando se emplearon como variables únicamente su composición en 11 ácidos grasos.

González (1997) también estudió la posibilidad de diferenciar el régimen alimenticio de cerdo Ibérico a partir de las características NIRS de su grasa. Para ello se utilizaron un total de 70 animales divididos en 7 lotes con diferentes alimentación: 4 de ellos a base de pienso compuesto comercial y 3 mediante el sistema tradicional de montanera. De ellos se obtuvieron muestras de tejido graso subcutáneo y de grasa perirrenal, las cuales se sometieron a dos diferentes tratamientos o formas de presentación (picada y fundida) previo a su análisis NIRS.

Se obtuvieron los espectros NIR de dichas muestras, tanto grasa subcutánea y perirrenal picada y fundida, en el rango de 1.100 a 2.500 nm mediante un instrumento IA 500 de Bran+Luebbe, estudiando la influencia de un tratamiento MSC de corrección de *scatter*. El análisis cualitativo de los espectros NIR se realizó mediante un análisis cluster (no supervisado).

Los resultados del análisis cluster resultaron claramente diferentes dependiendo del tipo de grasa, presentación y/o pretratamiento espectral considerado. Sin embargo, tras el análisis de los diferentes dendogramas obtenidos, en todos los casos se aprecian clasificaciones incorrectas de uno o varios lotes de animales alimentados en montanera en los lotes de animales alimentados en pienso o viceversa.

A modo de conclusión de la revisión bibliográfica de aplicaciones NIR en grasas y aceites, indicar que diversos trabajos mostraban la viabilidad de la tecnología NIRS para la determinación del contenido en ácidos grasos de grasa fundida de cerdo Ibérico (De Pedro *et al.*, 1992; González, 1997). Sin embargo, la precisión y exactitud de las ecuaciones de predicción del contenido en ácidos grasos obtenidas por diferentes autores para muestras de grasa fundida o tejido adiposo animal varían de forma

importante dependiendo de los trabajos considerados (De Pedro *et al.*, 1992; González, 1997; Ripoche y Guillard, 2001; Pérez-Marín *et al.*, 2002; Garrido *et al.*, 2002).

Teniendo en cuenta las numerosas fuentes de error asociadas al análisis NIRS y los distintos procedimientos de desarrollo de calibraciones NIRS, distintos instrumentos, diferente número y forma de presentación de muestras, distinto error del método de referencia y de la señal espectroscópica NIR y el diferente pretratamiento del dato espectral de estos trabajos, es difícil el concluir con exactitud sobre el factor o los factores responsables de dichas diferencias. Sin embargo, es de gran importancia el minimizar o eliminar estas fuentes de error y por tanto el disminuir los errores de predicción tanto en grasa fundida como en tejido adiposo.

Además de para la predicción del contenido en ácidos grasos, la información espectral NIR de grasas y aceites puede ser empleada para estudios de autenticación, clasificación o evaluación del grado de pureza. En concreto, diversos trabajos han mostrado las posibilidades de aplicación de los datos espectrales NIR para la evaluación del régimen alimenticio en cerdos Ibéricos (De Pedro *et al.*, 1992; González, 1997). Sin embargo, estos resultados se obtuvieron a partir de muestras procedentes de lotes de animales incluidos en proyectos de investigación y bajo diferentes algoritmos de clasificación. Por ello es necesario contrastar dichos resultados a partir de muestras de lotes de animales bajo condiciones reales (no diseñadas en proyectos de investigación) y comparar los resultados obtenidos mediante distintos métodos de clasificación.

La clonación de instrumentos y la transferencia de calibraciones NIRS, utilizando muestras sólidas no selladas es posible y es posible que, asimismo, pudiera ser de aplicación en el caso de la grasa fundida de cerdo Ibérico, si bien no se ha recogido en la bibliografía ningún trabajo en el que se desarrolle esta clonación y transferencia para muestras líquidas.

3 MATERIAL Y MÉTODOS

3.1 OPTIMIZACIÓN DE LA METODOLOGÍA PARA LA DETERMINACIÓN DE ÁCIDOS GRASOS EN CANALES DE CERDO IBÉRICO

3.1.1 ESTUDIO DE LOCALIZACIÓN ANATÓMICA DE TOMA DE MUESTRAS

3.1.1.1 Muestras de tejido adiposo empleadas

Para la evaluación de la posible existencia de diferencias en composición de ácidos grasos entre muestras de grasa obtenidas en diferentes localizaciones anatómicas en la canal, se tomaron muestras de grasa subcutánea de 30 canales de cerdos Ibéricos de tres zonas diferentes (Figura 3.1):

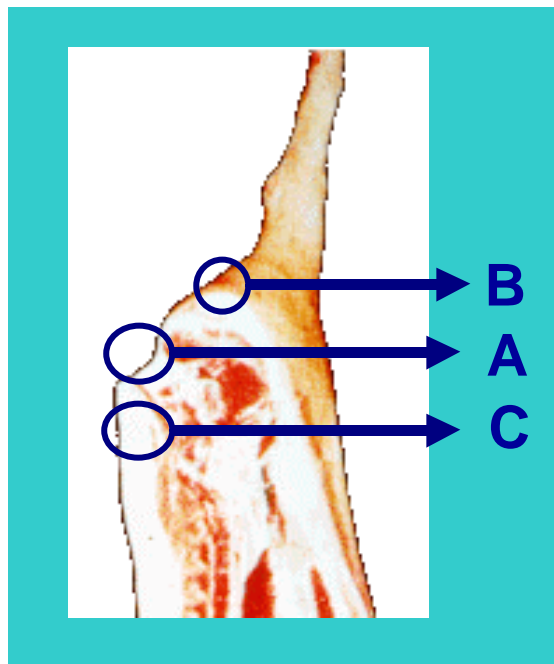
- A) En la inserción del rabo con el jamón (zona donde en el momento de realización de este estudio obtenía las muestras el Instituto de Tecnología Agroalimentaria de la Junta de Extremadura). El peso aproximado de esta muestra fue de 100 g.

- B) En el jamón, entre la inserción del rabo y la corva (zona donde obtenían las muestras algunas industrias del sector). El peso aproximado de esta muestra fue de 50 g.

- C) En la zona de inserción del rabo con la zona coxal (zona donde recomendaba la toma de muestras la Asociación Interprofesional del cerdo Ibérico, ASICI). El peso aproximado de esta muestra fue de 200 g.

Cada muestra de tejido adiposo de la canal así tomada se envasó en bolsas de plástico individuales previamente identificadas. El total de muestras obtenidas (90) se trasladó al Dpto. de Producción Animal (E.T.S.I.A.M.) de la Universidad de Córdoba (UCO) para su conservación bajo congelación (a -20°C) hasta su procesado.

Figura 3.1. Localización de las zonas de toma de muestras en la canal



3.1.1.2 Metodología de obtención de grasa líquida

Para llevar a cabo el análisis por cromatografía de gases de cada una de las muestras empleadas, es necesario la obtención de una muestra de grasa líquida a partir de cada una de las muestras de tejido adiposo tomadas de la canal. Entre los diferentes procedimientos de obtención de grasa líquida existentes, se utilizó el descrito por De Pedro *et al.* (1996) resumido a continuación, el cual realiza una fusión del tejido adiposo mediante calentamiento en horno microondas.

Tras la descongelación a temperatura ambiente de la muestra de tejido adiposo, se procede a la eliminación de los restos de piel, magro y la superficie externa de grasa y a su posterior troceado en piezas de reducido tamaño (aprox. 1 cm de arista).

Estas piezas así obtenidas se introducen en un recipiente de vidrio, sometiéndose a la acción de la radiación de microondas en un horno durante un período suficiente como para fundir la grasa (p. ej.: 3 minutos por cada 100 g de muestra a 700 W de potencia). De la grasa así fundida, y una vez retirados los restos de tejido subcutáneo sobrenadante, se obtiene la muestra líquida correspondiente que fue utilizada para el análisis por cromatografía de gases del presente estudio.

Este proceso de fusión se llevó a cabo, en el Dpto. de Producción Animal (E.T.S.I.A.M.) de la UCO, con el total de las muestras de tejido adiposo reseñadas en el apartado anterior, por lo que se obtuvieron 90 muestras de grasa líquida, correspondiente cada una de ellas a su respectiva muestra de tejido adiposo.

3.1.1.3 Determinación de ácidos grasos por cromatografía de gases

Las muestras de grasa líquida obtenidas se enviaron, bajo congelación, al Laboratorio Agroalimentario de la Junta de Andalucía en Córdoba para el análisis del contenido en ácidos grasos. El transporte se realizó, mediante neveras portátiles, utilizando tubos de polietileno conteniendo las muestras de grasa.

La composición de ácidos grasos se determinó mediante el análisis por cromatografía en fase gaseosa con columna capilar de los ésteres metílicos de los ácidos grasos, de acuerdo con la norma UNE 5508 (1996). La preparación de los ésteres metílicos se realizó siguiendo, en líneas generales, el procedimiento de metilación en frío indicado en el apartado 5 del método IUPAC nº 2301 (1987) el cual se resume a continuación.

Se disuelve una pequeña cantidad de grasa en hexano, se añade una disolución de KOH en metanol, se agita vigorosamente durante varios segundos y se deja decantar, tomando de la fase orgánica una alícuota para su análisis.

Todas las características de los métodos cromatográficos (equipos, columnas, reactivos, condiciones de operación y mantenimiento) empleados por el laboratorio para la separación y cuantificación de los ésteres metílicos, se encontraban dentro de las recomendaciones dadas por la norma UNE 5508 (1996).

En concreto se utilizó un cromatógrafo HP 5890 equipado con inyector automático y con una columna RTX-2330 (60 m de longitud, 0,25 mm de diámetro y 0,2 μm de espesor) con una composición al 10% de cianopropilfenil y un 90% de biscianopropilpolisiloxano. Las condiciones de trabajo del horno fueron de una temperatura inicial de 165°C durante 19 minutos, un incremento de temperatura de 5°C/min durante 7 minutos y una temperatura final de 200°C durante 2 minutos.

La cuantificación de los ácidos grasos se realizó a partir del cálculo del área de cada uno de los picos de los ésteres metílicos de los ácidos grasos considerados (C14:0, C16:0, C16:1, C17:0, C17:1, C18:0, C18:1, C18:2, C18:3, C18:4, C20:0, C20:1), teniendo en cuenta que la suma de las áreas cromatográficas fuese del 100% y empleando para ello los factores de respuesta obtenidos a partir de disoluciones de un patrón interno.

Sobre los datos aportados se aplicó un análisis general de varianza, utilizando un test Duncan para la determinación de diferencias de medias entre grupos, empleando el procedimiento PROC ANOVA del paquete estadístico SAS (1990).

3.1.2 ESTUDIO DEL NÚMERO MÍNIMO DE MUESTRAS DE GRASA REPRESENTATIVAS DE LA MEDIA DE UN LOTE

En el sector del cerdo Ibérico existe en general un desconocimiento acerca del número de animales a muestrear dentro de un lote con la misma alimentación destinada a sacrificio, con el fin de obtener una única muestra de grasa líquida representativa del lote para analizar por cromatografía de gases. Por este motivo se diseñó un estudio específico en el que se trata de determinar el número mínimo de muestras de grasa a tomar en un lote experimental para obtener una muestra de grasa líquida media de dicho lote.

Para ello se utilizaron 6 lotes de cerdos Ibéricos sometidos a diferente alimentación durante la etapa de cebo. Se obtuvo una muestra de grasa subcutánea de la canal de cada animal de los lotes considerados, en la zona de inserción del rabo con la zona coxal (zona C mencionada en el Apartado 3.1.1.1).

Cada muestra de grasa subcutánea fue fundida individualmente mediante la metodología ya descrita de fusión en horno microondas (De Pedro *et al.*, 1996). Posteriormente se procedió a la determinación del contenido en ácidos grasos de cada muestra mediante cromatografía de gases en el Laboratorio Agroalimentario de la Junta de Andalucía de Córdoba mediante el protocolo descrito en el apartado anterior.

En este estudio únicamente se consideraron aquellos ácidos grasos (C16:0, C18:0, C18:1, C18:2) que son empleados, en la evaluación de la alimentación de los cerdos Ibéricos por las diferentes Denominaciones de Origen del Sector y por el MAPA, mediante el contrato tipo homologado de compraventa de cerdos Ibéricos con destino a sacrificio.

Mediante la generación de números aleatorios, se seleccionaron diferentes porcentajes de animales a muestrear dentro de cada lote. Con el valor medio de las composiciones individuales de estos animales seleccionados se obtuvo una muestra media ficticia la cual fue comparada con la media total del lote.

Se ensayaron 4 porcentajes diferentes (10, 20, 30 y 40%) de animales a considerar para la obtención de la grasa media de un determinado lote. Mediante un programa de generación de números al azar, se obtuvieron, para cada uno de los 4 porcentajes ensayados, la identificación de los animales de cada partida que deberían constituir la muestra media, representativa de cada partida. Dicha identificación correspondía al número de orden de la canal en la línea de sacrificio, dentro de cada partida. Para conocer la posible influencia de la toma de muestras, en cada uno de estos porcentajes estudiados, se consideraron 5 repeticiones o series de toma de muestras al azar.

De cada serie de muestras y para cada porcentaje de animales de los distintos lotes empleados, se determinó la media de composición de los 4 ácidos grasos considerados, así como la clasificación de acuerdo con los criterios analíticos incluidos en el contrato tipo homologado de compra-venta de cerdos Ibéricos para la campaña 1998/1999 (BOE, 1998). Para la generación de número aleatorios y para la obtención de los valores medios de los ácidos grasos considerados se empleó la hoja de cálculo Microsoft Excel.

3.1.3 ENSAYO COLABORATIVO

El ensayo colaborativo se efectuó para conocer el error intra e interlaboratorial existente en la determinación del contenido de ácidos grasos en grasa de cerdo Ibérico mediante cromatografía de gases. En este ensayo, coordinado por el Dpto. de Producción Animal de la Universidad de Córdoba, participaron seis laboratorios donde se realizan la práctica totalidad de los análisis de grasa de cerdo Ibérico para industriales y ganaderos del Sector. Los laboratorios participantes fueron:

- Laboratorio Agroalimentario de la Consejería de Agricultura y Pesca de la Junta de Andalucía, en Córdoba.
- Instituto de Tecnología Agroalimentaria de la Consejería de Agricultura y Medio Ambiente de la Junta de Extremadura, en Badajoz.
- Estación Tecnológica de la Carne de la Consejería de Agricultura y Ganadería de la Junta de Castilla y León, en Guijuelo (Salamanca).
- Laboratorio de la Asociación Interprofesional del cerdo Ibérico (ASICI), en Zafra (Badajoz).
- Laboratorio Arbitral Agroalimentario del Ministerio de Agricultura Pesca y Alimentación, en Madrid.
- Dpto. de Química Analítica, Nutrición y Bromatología de la Universidad de Salamanca, en Salamanca.

3.1.3.1 Preparación y análisis de muestras

Para este ensayo se emplearon muestras de grasa subcutánea de 30 canales de cerdos Ibéricos finalizados con un régimen alimenticio a base de pienso. Las muestras se tomaron de la zona de inserción del rabo en la región coxal (zona C mencionada en el Apartado 3.1.1.1). Tras eliminar los restos de piel y magro de cada muestra de grasa, estas fueron divididas en 2 submuestras. Dichas submuestras de tejido adiposo fueron procesadas de acuerdo con 2 metodologías diferentes de extracción de grasa líquida:

- Fusión mediante microondas: ya descrita en el Apartado 3.1.1.2. Este proceso se realizó sobre las 30 muestras de tejido adiposo en el Dpto. de Producción Animal (E.T.S.I.A.M.) de la Universidad de Córdoba.
- Extracción a temperatura ambiente con disolvente: Un máximo de 5 g de muestra son picados y homogeneizados antes de introducirlos en bolsas estériles. Una vez en las bolsas se añade un disolvente (éter dietílico) para la extracción de la grasa. La bolsa, con la grasa picada y el disolvente se introducen en un homogeneizador de palas y se activa durante varios minutos. Transcurrido el tiempo de extracción, el contenido de la bolsa se filtra para separar el residuo sólido. El disolvente contenido en la fase líquida se evapora, bajo vacío, en un evaporador rotativo a una temperatura de 50°C. El extracto líquido restante tras la evaporación constituye la muestra de grasa empleada en el análisis por cromatografía de gases. Este proceso se realizó en el laboratorio de ASICI, a las 30 muestras de tejido adiposo mencionadas anteriormente.

Se dispuso por tanto de un total de 60 muestras de grasa fundida, 30 obtenidas mediante fusión con microondas y 30 obtenidas mediante extracción con disolvente.

Este conjunto total de 60 muestras fueron las empleadas para llevar a cabo el estudio colaborativo. Para ello, cada uno de los 6 laboratorios participantes recibió una submuestra de cada una de las 60 muestras, mediante mensajería urgente y a temperatura de congelación con la recomendación de realizar la determinación del contenido en ácidos grasos por duplicado, para poder evaluar la repetibilidad del método de análisis.

La composición de ácidos grasos se determinó en cada laboratorio participante de acuerdo con la metodología expuesta en el apartado 3.1.1.2 y las recomendaciones dadas por la norma UNE 5508 (1996).

3.1.3.2 Tratamiento estadístico de resultados

El tratamiento estadístico de los datos obtenidos en este estudio colaborativo se realizó de acuerdo con la norma ISO 5725-2 (1994) o la correspondiente norma

española UNE 82009-2 (1999), las cuales detallan la metodología a seguir para la determinación de la repetibilidad y reproducibilidad de un método de medida estándar. En ellas se especifican los criterios empleados para el rechazo de resultados anómalos y los procedimientos estadísticos empleados en el cálculo del error intralaboratorio (repetibilidad) e interlaboratorios (reproducibilidad).

El caso general de un ensayo colaborativo es aquel en el que participan p laboratorios denominados i ($i=1, 2, \dots, p$), cada uno de los cuales analiza q muestras, denominadas j ($j=1, 2, \dots, q$) con n replicas por muestra, denominadas k ($k=1, 2, \dots, n$) dando un total de $(p \times q \times n)$ resultados. En este caso, participaron 6 laboratorios ($p=6$) analizando cada uno de ellos 60 muestras ($q=60$) por duplicado ($n=2$).

Los estadísticos básicos empleados en el ensayo colaborativo fueron:

- Valor medio de las n repeticiones obtenidas por laboratorio y por muestra. Su fórmula es:

$$\bar{y}_{ij} = \frac{\sum_{k=1}^{n_{ij}} y_{ijk}}{n_{ij}}$$

- Desviación típica de las n repeticiones de cada muestra obtenidas por cada laboratorio (s_{ij}). Su fórmula es:

$$s_{ij} = \sqrt{\frac{\sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2}{n_{ij} - 1}}$$

donde n_{ij} es el número de réplicas de una muestra j analizada por un laboratorio i , y_{ijk} es el resultado de cada una de las réplicas ($k=1, \dots, n_{ij}$) y p_j es el número de laboratorios implicados en el ensayo colaborativo.

A partir de estos estadísticos, en el análisis se calcula, para cada muestra y de forma independiente, los valores de los siguientes estadísticos:

- Media general (\bar{y}_j): valor medio entre laboratorios para cada una de las muestras. Su fórmula es:

$$y_j = \frac{\sum_{i=1}^p n_{ij} \bar{y}_{ij}}{\sum_{i=1}^p n_{ij}}$$

- Varianza de la Repetibilidad (s_{rj}^2): varianza del error intralaboratorio para cada una de las muestras:

$$s_{rj}^2 = \frac{\sum_{i=1}^p (n_{ij} - 1) s_{ij}^2}{\sum_{i=1}^p (n_{ij} - 1)}$$

En el caso de que se realicen 2 réplicas por muestra en los laboratorios, es decir, análisis por duplicado ($n_{ij}=n=2$), la expresión de la varianza del error entre laboratorios se simplifica:

$$s_{rj}^2 = \frac{\sum_{i=1}^p (y_{ij1} - y_{ij2})^2}{2p}$$

- Varianza entre laboratorios (s_{Lj}^2): varianza del error existente entre laboratorios sin incluir el error intralaboratorio. Su fórmula es:

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{n_j}$$

donde

$$s_{dj}^2 = \frac{\sum_{i=1}^p n_{ij} (\bar{y}_{ij} - \bar{y}_j)^2}{p-1}$$

En el caso de que se realicen 2 réplicas por muestra en los laboratorios, es decir, análisis por duplicado ($n_{ij}=n=2$), la expresión de la varianza del error entre laboratorios se simplifica a:

$$\bar{n}_j = \frac{1}{p-1} \left[\sum_{i=1}^p n_{ij} - \frac{\sum_{i=1}^p n_{ij}^2}{\sum_{i=1}^p n_{ij}} \right]$$

$$s_{Lj}^2 = \frac{\sum_{i=1}^p (\bar{y}_{ij} - \bar{y}_j)^2}{p-1} - \frac{s_{rj}^2}{2}$$

- Varianza de la Reproducibilidad (s_{Rj}^2): varianza del error existente entre laboratorios incluido el error intralaboratorio. Su fórmula es:

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2$$

Como etapa previa al cálculo de dichos estadísticos, se ha de eliminar aquellos resultados detectados como anómalos. Para ello, la norma ISO 5725-2 (1994) recomienda 2 procedimientos: una primera aproximación mediante técnicas de representación gráfica y una segunda mediante test numéricos.

Como técnicas gráficas se emplean la representación de los estadísticos h y k de Mandel.

- h de Mandel: nos informa acerca de la consistencia entre laboratorios. Se calcula dividiendo, para cada laboratorio, el residual entre el valor medio (media de un laboratorio i menos la media general) por la desviación típica entre los valores medios de los laboratorios:

$$h_{ij} = \frac{\bar{y}_{ij} - \bar{y}_j}{\sqrt{\frac{\sum_{i=1}^p (\bar{y}_{ij} - \bar{y}_j)^2}{p_j - 1}}}$$

- k de Mandel: nos informa acerca de la consistencia dentro de cada laboratorio. Se calcula dividiendo, para cada laboratorio y por muestra, la desviación típica de las repeticiones por la desviación típica de las repeticiones ponderada:

$$k_{ij} = \frac{s_{ij}}{\sqrt{\frac{\sum s_{ij}^2}{p_j}}}$$

Un resultado se considera anómalo h o k cuando sobrepasa los valores tabulados (ISO, 1994) al nivel del 1%. Si es inferior a los valores tabulados al nivel del 5% pero superior a los del nivel del 1%, el resultado es considerado como dudoso.

Como test numéricos se aplicaron los test de Cochran (calcula la variabilidad existente dentro de cada laboratorio) y el test de Grubbs sencillo y doble (calcula la variabilidad existente entre los laboratorios). Mediante el test de Cochran se determinan muestras y/o laboratorios con varianzas significativamente elevadas, es decir, aquellas/os cuyas diferencias entre resultados de réplicas son excepcionalmente altas. Mediante el test de Grubbs se determinan muestras y/o laboratorios con valores extremos inconsistentes con una distribución normal de los resultados, es decir aquellos que aportan valores que son significativamente diferentes a los obtenidos por el resto.

- Test de Cochran: Se calcula, para cada una de las muestras de forma independiente, el estadístico de Cochran (C_j), el cual viene dado por la fórmula:

$$C_j = \frac{s_{ij}^2 \max}{\sum_{i=1}^{p_j} s_{ij}^2}$$

donde $s_{ij}^2 \max$ es el mayor valor de s_{ij}^2 por muestra. Si el resultado es superior al valor tabulado al nivel de 1%, una muestra es considerada como anómalo (outlier). En el caso de que un determinado resultado aportado por un laboratorio para una determinada muestra sea detectado como anómalo, se ha de repetir el test de Cochran eliminando dicho valor. Si se detectan varios anómalos para muestras diferentes del mismo laboratorio, esto indicaría que la varianza dentro de dicho laboratorio es anormalmente alta por lo que todos los datos aportados por él podrían llegar a ser eliminados.

- Test de Grubbs: Se aplican 2 test diferentes para detectar la existencia de un único anómalo o bien de 2 observaciones anómalas. Para su cálculo es recomendable la ordenación de los valores medios (\bar{y}_{ij}) en orden ascendente o descendente en cada una de las muestras.
 - 1 observación anómala: Se calcula el estadístico de Grubbs (G_I) para evaluar si los valores medios máximos o mínimos de cada muestra son anómalos.

$$G_1max = \frac{(\bar{y}_{ij}max - \bar{y}_j)}{\sqrt{\frac{\sum_{i=1}^p (\bar{y}_{ij} - \bar{y}_j)^2}{p-1}}}$$

$$G_1min = \frac{(\bar{y}_{ij}min - \bar{y}_j)}{\sqrt{\frac{\sum_{i=1}^p (\bar{y}_{ij} - \bar{y}_j)^2}{p-1}}}$$

donde $y_{ij}max$ e $y_{ij}min$ son los valores medios máximos y mínimos respectivamente en cada una de las muestras. Un resultado se considera anómalo cuando el valor de G_1max o G_1min es mayor que el valor tabulado (ISO, 1994) al nivel del 1%. Si sobrepasa los valores tabulados al nivel del 5%, el resultado es considerado como dudoso.

- 2 observaciones anómalas: Se calcula el estadístico de Grubbs (G_2) para evaluar si los 2 valores medios más altos o más bajos de cada muestra son anómalos.

$$G_2max = \frac{s_{p-1,p}^2}{s_o^2} = \frac{\sum_{i=1}^{p-2} (\bar{y}_{ij} - \bar{y}_{j(p-1,p)})^2}{\sum_{i=1}^p (\bar{y}_{ij} - \bar{y}_j)^2}$$

$$G_2min = \frac{s_{1,2}^2}{s_o^2} = \frac{\sum_{i=3}^p (\bar{y}_{ij} - \bar{y}_{j(1,2)})^2}{\sum_{i=1}^p (\bar{y}_{ij} - \bar{y}_j)^2}$$

donde $y_{j(p-1,p)}$ e $y_{j(1,2)}$ son los 2 valores medios máximos y mínimos respectivamente en cada una de las muestras. Un resultado se considera anómalo cuando el valor de G_2max o G_2min es inferior que el valor tabulado (ISO, 1994) al nivel del 1%. Si es inferior a los valores tabulados al nivel del 5% pero superior a los del nivel del 1%, el resultado es considerado como dudoso.

Una vez detectados los resultados anómalos, dichos datos son eliminados y con el resto se calculan los estadísticos de varianza y desviación típica de la

repetibilidad (s_{rj}^2 ; s_{rj}), varianza y desviación típica entre laboratorios (s_{Lj}^2 ; s_{Lj}) y varianza y desviación típica de la reproducibilidad (s_{Rj}^2 ; s_{Rj}) en cada una de las muestras.

A continuación se evalúa la relación o dependencia de dichas desviaciones típicas frente a los valores medios de las muestras. Si no existiera ningún tipo de dependencia (situación más probable), el valor final de desviación típica de la repetibilidad y reproducibilidad se calcula como el valor medio de todas las muestras. Es decir:

$$\text{Desviación típica de la repetibilidad} = s_r = \frac{\sum_{j=1}^q s_{rj}}{q}$$

$$\text{Desviación típica de la reproducibilidad} = s_R = \frac{\sum_{j=1}^q s_{Rj}}{q}$$

De acuerdo con la norma AOAC (1999) para la realización de estudios colaborativos, se pueden definir los parámetros de repetibilidad y reproducibilidad como múltiplos de dichas medidas de precisión. Las definiciones vienen dadas en forma de intervalos de predicción e indican que la máxima diferencia absoluta entre 2 resultados, y solo 2, obtenidos sobre la misma muestra ya sea en un mismo laboratorio (repetibilidad) o en 2 laboratorios diferentes (reproducibilidad) no serán superiores al valor de repetibilidad (r) o reproducibilidad (R) en un 95% de los casos. Para ello se asume una distribución normal de los resultados y un nivel de probabilidad de 95%. Las fórmulas para los valores de repetibilidad (r) y reproducibilidad (R) son, de acuerdo con la definición de AOAC (1999), las que se indican a continuación:

$$r = \text{repetibilidad} = \sqrt{2} \times t(N(a-1), 95\%) \times s_r$$

$$R = \text{reproducibilidad} = \sqrt{2} \times t(N(a-1), 95\%) \times s_R$$

donde $t(N(a-1), 95\%)$ es el valor de la t de Student para un nivel de confianza del 95% y $N(a-1)$ grados de libertad, N el número de muestras y a el de replicas de la misma muestra. En la mayoría de los casos, el producto $N(a-1)$ suele ser elevado (en este caso

60) de forma que se puede aproximar el valor de la t de Student para un intervalo de confianza del 95% a 1,96 ó 2. De esta forma, las expresiones anteriores quedarían:

$$r = \text{repetibilidad} = \sqrt{2} \times 2 \times s_r = 2,8 \times s_r$$

$$R = \text{reproducibilidad} = \sqrt{2} \times 2 \times s_R = 2,8 \times s_R$$

Otro parámetro estadístico importante es la desviación estándar relativa de la repetibilidad (RSD_r) y de la reproducibilidad (RSD_R). Aportan información ya que normalmente son independientes de la concentración del analito en un rango razonable. Ello facilita la comparación de las variabilidades a diferentes concentraciones o entre varios analitos diferentes, en este caso, ácidos grasos. Sus fórmulas son:

$$RSD_r (\%) = \frac{s_r}{\bar{y}} \times 100$$

$$RSD_R (\%) = \frac{s_R}{\bar{y}} \times 100$$

donde \bar{y} es la media global de cada ácido graso para todas las muestras y laboratorios.

En el tratamiento de datos aportados por los laboratorios participantes y en cálculo de los estadísticos expuestos con anterioridad, se empleó la hoja de cálculo Microsoft Excel.

Para la comparación de los valores medios de cada ácido graso obtenidos por los diferentes laboratorios y los métodos propuestos de extracción de grasa del tejido adiposo, se empleó un análisis estadístico de la varianza (ANOVA) junto con un test Duncan de comparación de medias mediante el procedimiento PROC ANOVA del programa SAS (1990).

3.2 PUESTA A PUNTO DE METODOLOGÍAS PARA OBTENCIÓN DE INFORMACIÓN ESPECTROSCÓPICA NIR EN GRASA

3.2.1 MODALIDAD DE DOBLE TRANSMISIÓN

3.2.1.1 Muestras

Las muestras de grasa subcutánea empleadas en este estudio se tomaron de acuerdo con la sistemática descrita en el Apartado 3.1.3.1 a partir de canales de cerdos producidos al amparo de la D.O. “Jamón de Huelva”. Dichas muestras fueron transportadas al laboratorio de Dpto. de Producción Animal, bajo congelación, en neveras isotérmicas.

Una vez en dicho laboratorio, las muestras de tejido adiposo fueron descongeladas hasta temperatura ambiente en el laboratorio procediéndose a continuación a su fusión en horno microondas de acuerdo con la metodología descrita en el Apartado 3.1.1.2. Tras eliminar los restos de tejido adiposo sobrenadante, se tomaron varias alícuotas de grasa líquida las cuales fueron envasadas en viales de plástico y conservadas en congelación (-20°C) hasta la obtención de sus espectros.

3.2.1.2 Análisis NIRS

Los espectros de grasa líquida fueron obtenidos mediante la modalidad de doble transmisión empleando para ello una cápsula circular de 0,1 mm de paso óptico con fondo de aluminio como material reflectante (nº catalogo IH-0345). Dadas las

características comentadas de esta modalidad de análisis NIRS en la cual la radiación atraviesa dos veces la muestra tras reflejarse en el fondo de la cápsula, el paso óptico efectivo en esta modalidad es el doble del paso óptico del que disponga la cápsula de análisis. Así, en este caso el paso óptico efectivo fue de 0,2 mm.

Las muestras fueron analizadas mediante un espectrofotómetro Foss NIRSystems 6500 SY I equipado con un módulo de giro. Los espectros se obtuvieron, con un intervalo de 2 nm, en el rango comprendido entre 400 y 2.500 nm, empleando para ello el programa ISI NIRS3 ver. 2.05 (Infrasoft International, Port Matilda, PA, USA). Se obtuvo 1 espectro de la cerámica de referencia o patrón interno antes y después del análisis NIRS de cada muestra. A su vez, cada espectro de referencia era el resultado de mediar 16 barridos de la radiación (scans) sobre la cerámica. En el caso de la muestra a analizar, sus datos espectrales se correspondían con el espectro medio de 32 barridos de la radiación (scans) sobre la muestra.

3.2.1.3 Cálculo de la repetibilidad espectral

Se empleó el estadístico *RMS*, incluido en el programa MANAGE de ISI NIRS 3 ver. 3.11. (Shenk y Westerhaus, 1995b), para calcular las similitudes o diferencias existentes en valores de absorbancia entre varios espectros. Este programa proporciona los valores *MEAN* y *STD* del estadístico *RMS* para un grupo de submuestras (espectros) así como los valores individuales de *RMS* para cada una de los espectros de dicho grupo. El valor del estadístico *RMS* para una única submuestra *j* (*RMS_j*) viene dado por la ecuación:

$$RMS_j = 10^6 \times \sqrt{\frac{\sum_{i=1}^n D_{ij}^2}{n}}; \quad D_{ij} = y_{ij} - \bar{y}_i$$

$y_{ij} = \log(1/R)$ a \mathbf{I}_i para una submuestra (espectro) *j*.

$\bar{y}_i = \log(1/R)$ a \mathbf{I}_i para el espectro medio de *N* submuestras de una muestra.

$n = 400, 402, \dots, 2500 = 1050$ datos espectrales.

El hecho de introducir un factor multiplicativo de 10^6 en la fórmula anterior es para obtener un rango numérico de dicho estadístico más manejable.

Los valores *MEAN* y *STD* del estadístico *RMS* para un grupo de espectros se calculan de acuerdo con las siguientes fórmulas (Rosales, 1993):

$$MEAN = \sqrt{\frac{\sum_{j=1}^N (RMS_j)^2}{N}} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^N (y_{ij} - \bar{y}_i)^2}{nN}} \quad [1]$$

$$STD = \sqrt{\frac{\sum_{j=1}^N (RMS_j)^2}{N-1}} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^N (y_{ij} - \bar{y}_i)^2}{n(N-1)}} \quad [2]$$

Teniendo en cuenta [1] y [2], la relación existente entre los valores *MEAN* y *STD* del estadístico *RMS* viene definida por la siguiente expresión:

$$STD = \sqrt{\frac{N}{N-1}} \times MEAN \quad [3]$$

En el caso de 2 submuestras ($N=2$), la anterior expresión quedaría simplificada a:

$$STD = \sqrt{2} \times MEAN \quad [4]$$

Estas dos últimas fórmulas resultan muy útiles si se tiene en cuenta que ISI no realiza correctamente el cálculo del estadístico *STD* en versiones posteriores a la 2.05. Al ser el cálculo del estadístico *MEAN* correcto en todas las versiones, es posible calcular el valor del *STD* a partir de [3] o [4] según proceda.

El cálculo del estadístico *STD* resulta muy útil ya que permite su uso en el análisis de rutina para evaluar la repetibilidad espectral. De acuerdo con Rosales (1993), este valor de *STD* es una expresión de una varianza del error, con una distribución muestral de una χ^2 . Por lo tanto, a la hora de establecer un criterio de decisión sobre cual ha de ser el valor de *STD* límite por encima del cual las submuestras podrían considerarse no representativas del material analizado, se ha de emplear la siguiente expresión:

$$STD_{limite} = 1,036 \times \sqrt{\frac{\sum_{k=1}^m STD_k^2}{m}} = 1,036 \times \sqrt{STD^2}$$

siendo *m* el número de muestras analizadas.

Para estimar el valor de \overline{STD} , Rosales (1993) indica que existen dos aproximaciones, si bien muestra que ambas conducen a un valor estimado similar.

Para evaluar la repetibilidad espectral en grasa fundida de cerdo Ibérico analizada mediante doble transmisión eliminando el efecto del relleno de cápsulas, se obtuvieron 10 espectros de una misma muestra, sin vaciar el contenido de la cápsula. Dicho ensayo se realizó tanto con la muestra de grasa en estado pastoso (24°C) como en estado líquido (40°C) para poder observar la influencia de la temperatura en la repetibilidad de la señal. Una vez obtenida la información espectral, se calcularon los valores de *RMS* entre diferentes combinaciones de grupos de 2 espectros. Estos cálculos se realizaron de forma independiente para 30 comparaciones de espectros de grasa líquida, 30 de grasa en estado pastoso y 30 combinaciones de espectros en ambos estados. Posteriormente se estimó el valor de \overline{STD} y $STD_{límite}$.

Otra magnitud de error también de interés en el cálculo de la repetibilidad espectral intramuestra, esto es, aquella debida a las diferencias entre submuestras de la misma muestra, provocadas por el relleno de la cápsula. Para evaluar dicho error se obtuvieron los espectros de 10 submuestras de una misma muestra, mediante doble transmisión. A partir de ellos se calcularon los valores de *RMS* y los estadísticos *MEAN* y *STD* entre 20 combinaciones de pares de espectros pertenecientes a diferentes submuestras. Posteriormente se estimó el valor de \overline{STD} y $STD_{límite}$.

Un factor operacional que podría tener gran influencia en la repetibilidad de la medida espectroscópica mediante esta modalidad de análisis NIRS es la presencia de burbujas de aire en el interior de la muestra, tanto en estado líquido como en estado pastoso.

Con el fin de evaluar el efecto de dicho factor, se realizó un ensayo en el que, deliberadamente, se provocó la presencia de burbujas en el interior de la cápsula. Se obtuvieron 9 espectros de una misma muestra de grasa líquida de cerdo Ibérico de forma que 3 de ellos correspondían a cápsulas donde no se obtuvieron burbujas en su relleno, 5 a cápsulas con poca presencia de burbujas (entre 10 y 15) y 1 a una cápsula con gran cantidad de aire en su interior (aproximadamente, el 40% del volumen). A continuación se calcularon los valores de los estadísticos *RMS*, *MEAN* y *STD* para todas

las combinaciones posibles entre pares de espectros con diferente presencia de burbujas de aire en su interior.

3.2.2 MODALIDAD DE INTERACTANCIA-REFLECTANCIA.

3.2.2.1 Muestras

Las muestras de grasa subcutánea empleadas en el presente apartado, eran pertenecientes a canales de cerdos producidos al amparo de la D.O. “Jamón de Huelva” y se tomaron de acuerdo con la sistemática descrita en el apartado 3.1.3.1. Dichas muestras se transportaron al Dpto. de Producción Animal en neveras isotérmicas, bajo congelación, y se conservaron a -20°C hasta su procesado.

Una vez en dicho laboratorio, las muestras fueron descongeladas hasta temperatura ambiente y a continuación atemperadas a la temperatura fija existente en la sala de los equipos NIRS (aprox. 24°C). Previo al análisis espectroscópico NIR de cada muestra, se procedió al corte y eliminación de una capa de 1 cm de grosor en la superficie donde se iba a llevar a cabo la toma de los espectros para evitar, en la medida de lo posible, la presencia de grasa oxidada (enrancada). Además, se seleccionó aquella porción de muestra donde se podía apreciar de forma nítida la presencia de las 2 capas existentes en el tejido adiposo de cerdo Ibérico (externa e interna) para así poder evaluar las diferencias espectrales existentes entre ellas.

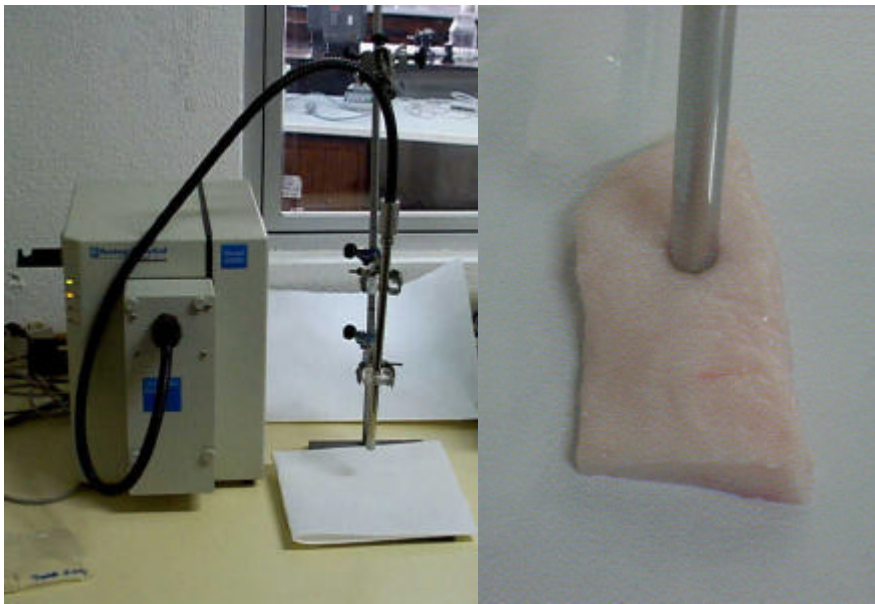
3.2.2.2 Análisis NIRS

Los espectros fueron obtenidos mediante la modalidad de interactancia-reflectancia mediante el uso de una sonda de fibra óptica. En esta modalidad, la fuente de la radiación NIR y los detectores se encuentran en la misma posición respecto a la muestra, incidiendo la radiación en la muestra con un ángulo determinado y reflejándose de ella con diferentes ángulos de emisión (Figura 3.2).

Las muestras fueron analizadas mediante un espectrofotómetro Foss NIRSystems 6500 SY II equipado con un módulo de sonda de fibra óptica (ref. nº NR-6775). Los espectros se obtuvieron, con un intervalo de 2 nm, en el rango comprendido

entre 800 y 2.200 nm de acuerdo con las recomendaciones del fabricante, empleando para ello el programa ISI 2 ver. 4.0 (Infrasoft International, Port Matilda, PA, USA). Se obtuvo 1 espectro de la referencia o patrón interno cada hora el cual era el resultado de mediar 12 barridos de la radiación (scans) sobre la cerámica. En el caso de la muestra a analizar, sus datos espectrales se correspondían con el espectro medio de 24 barridos de la radiación (scans) sobre la muestra.

Figura 3.2. Obtención de espectros de tejido subcutáneo de cerdo Ibérico mediante la modalidad de interactancia-reflectancia.



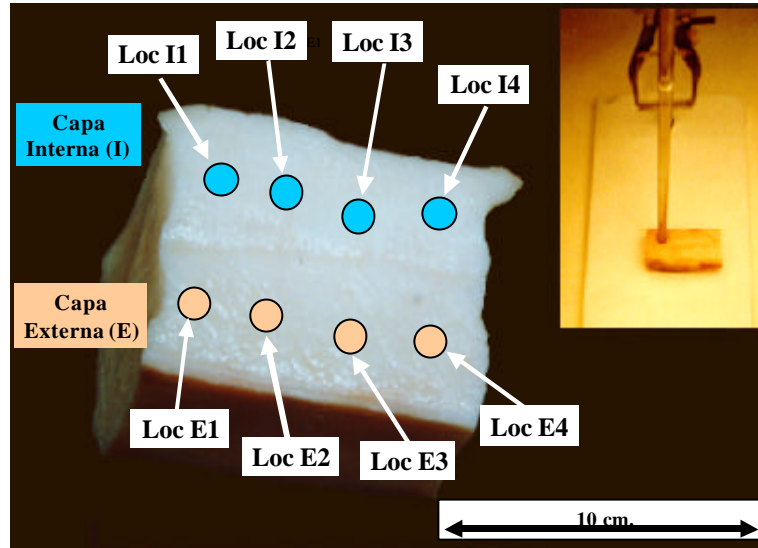
3.2.2.3 Cálculo de la repetibilidad espectral

Se evaluó el efecto de diversos factores sobre la repetibilidad de los datos espectrales obtenidos mediante la modalidad de análisis de interactancia-reflectancia. Estos factores fueron el número y localización de los espectros a tomar en una muestra de tejido adiposo y el efecto de luz ambiental sobre el espectro.

Para evaluar el número y localización de los espectros a tomar de cada muestra de tejido adiposo, utilizando la mencionada sonda de fibra óptica, se seleccionó una muestra de grasa subcutánea en la cual se pudiera apreciar de forma nítida la presencia de las 2 dos capas de tejido adiposo (externa e interna). En dicha muestra se obtuvieron espectros en 8 posiciones, 4 de ellas situadas en la capa externa y 4 en la capa interna de tejido adiposo (Figura 3.3). A su vez, para evaluar la repetibilidad de la

señal espectroscópica se obtuvieron 4 espectros consecutivos, sobre cada una de las localizaciones, sin mover de posición la sonda de fibra óptica.

Figura 3.3. Tejido adiposo con diferentes posiciones de la sonda de fibra óptica en la capa externa (E) e interna (I).



Una vez obtenida la información espectral para cada localización, se calcularon los valores de los estadísticos *RMS*, *MEAN* y *STD* (apartado 3.2.1.3) entre diferentes combinaciones de grupos de 2 espectros. Por una parte entre pares de espectros obtenidos en la misma localización de la muestra, evaluando con ello el error existente dentro de cada localización de la sonda. Por otra, entre pares de espectros tomados en diferentes localizaciones de una misma capa (externa o interna), estimando así el error de la medida espectral entre diferentes localizaciones de la misma capa grasa. Finalmente, se calcularon los valores de *RMS*, *MEAN* y *STD* entre pares de espectros obtenidos en localizaciones de distintas capas de tejido adiposo, evaluando así el error espectral existente entre espectros de diferentes capas de grasa subcutánea.

Un factor operacional que podría tener gran influencia en la repetibilidad de la medida espectroscópica mediante esta modalidad de análisis NIRS es la presencia de luz ambiental o luz procedente de alguna lámpara externa que pudiera alcanzar los detectores del instrumento a través de la sonda de fibra óptica (Williams y Norris, 1987a; Mitsumoto *et al.*, 1991).

Para evaluar el efecto de la luz ambiental sobre la repetibilidad espectral, se obtuvieron 4 espectros en 3 diferentes posiciones de una muestra (Loc A, Loc B y Loc

C) tanto protegiendo como sin proteger a la muestra mediante una caja negra que evitara, en la medida de lo posible, la presencia de luz procedente del exterior. A continuación se calcularon los valores de los estadísticos *RMS*, *MEAN* y *STD* para todas las combinaciones posibles de pares de espectros con ausencia o presencia de luz procedente del exterior.

Todas las transformaciones y tratamientos matemáticos de los datos espectroscópicos se llevaron a cabo mediante el paquete quimiométrico ISI NIRS 3 ver. 3.11. Las fórmulas empleadas en los cálculos de los estadísticos *RMS*, *MEAN*, *STD* y *STD_{límite}* son las mismas que las indicadas en el apartado 3.2.1.3. Únicamente se ha de tener en cuenta que el número de datos espectrales considerados fue menor ($n=700$) al emplear un intervalo de longitudes de onda más pequeño (800 a 2.200 nm).

3.2.3 INFLUENCIA DE LA CONGELACIÓN DE MUESTRAS DE GRASA DE CERDO IBÉRICO SOBRE SU INFORMACIÓN ESPECTRAL NIR.

3.2.3.1 Muestras

Para evaluar este factor operacional, posible fuente de error espectral NIR, se han utilizado muestras de tejido adiposo de cerdos ibéricos pertenecientes a 10 partidas sacrificadas durante la campaña 97/98. La toma de muestras y fusión del tejido adiposo mediante microondas se realizó de acuerdo con la metodología descrita en el apartado 3.1.3.1.

Se obtuvo una muestra de grasa líquida media representativa de cada una de los 10 lotes empleados en el estudio. A partir de cada una de dichas muestras líquidas se elaboraron 6 submuestras de 10 mililitros de cada una de los 10 lotes, las cuales fueron envasadas en tubos de ensayo de polietileno, convenientemente etiquetadas y conservadas bajo congelación a -20°C hasta su posterior análisis NIRS. Se dispuso por tanto de un total de 60 muestras de grasa líquida.

3.2.3.2 Análisis NIRS

Se evaluaron 2 factores de conservación que, a priori, podían afectar a la repetibilidad espectral NIR de la grasa fundida de cerdo Ibérico como son la influencia del tiempo de conservación de las muestras bajo congelación y la influencia del efecto de congelación-descongelación sobre la grasa fundida.

Para ello, se obtuvieron a lo largo de 6 meses (Mes 0 a Mes 5) los espectros NIR de las muestras reseñadas en el apartado anterior de acuerdo con el diseño descrito en la Tabla 3.1. De acuerdo con dicha tabla, cada una de las series de alícuotas analizadas, es decir, series de espectros obtenidos, se referencian como $T_{i,j}$ donde i corresponde al mes en el que el tubo de submuestra se abre por vez primera mientras que j indica el número de veces que el tubo de submuestra ha sufrido el proceso de congelación-descongelación.

Tabla 3.1. Diseño experimental del estudio de conservación de grasa*.

Mes Análisis	Nº espectros por nº veces de análisis de muestra tras congelación-descongelación						Total Análisis
	1ª vez	2ª vez	3ª vez	4ª vez	5ª vez	6ª vez	
MES 0	10 ($T_{0,0}$)						10
MES 1	10 ($T_{1,0}$)	10 ($T_{0,1}$)					20
MES 2	10 ($T_{2,0}$)	10 ($T_{1,1}$)	10 ($T_{0,2}$)				30
MES 3	10 ($T_{3,0}$)	10 ($T_{2,1}$)	10 ($T_{1,2}$)	10 ($T_{0,3}$)			40
MES 4	10 ($T_{4,0}$)	10 ($T_{3,1}$)	10 ($T_{2,2}$)	10 ($T_{1,3}$)	10 ($T_{0,4}$)		50
MES 5	10 ($T_{5,0}$)	10 ($T_{4,1}$)	10 ($T_{3,2}$)	10 ($T_{2,3}$)	10 ($T_{1,4}$)	10 ($T_{0,5}$)	60

$T_{i,j}$; i = mes en el que el tubo de submuestra del que procede se abre por primera vez; j = número de veces en que el tubo de submuestra del que procede ha sufrido el proceso de congelación-descongelación.

Al inicio de la experiencia (momento de fusión y preparación de muestras, $T_{0,0}$) se analizó una alícuota de grasa de una sola de las submuestras de cada muestra, obteniéndose así 10 espectros (uno por cada muestra) y conservándose de nuevo las muestras bajo congelación (-20°C). Transcurrido un mes, se volvieron a obtener espectros de una alícuota de las submuestras abiertas el primer mes ($T_{0,1}$) y además se obtuvo el espectro NIR de 10 nuevas submuestras ($T_{1,0}$), una por cada muestra, obteniéndose el Mes 1 un total de 20 espectros.

El segundo mes se analizaron de nuevo alícuotas de las veinte submuestras analizadas el mes anterior ($T_{0,2}$ y $T_{1,1}$) y además las contenidas en otros diez tubos abiertos por primera vez ($T_{2,0}$). Así sucesivamente, hasta el quinto mes de análisis en el que se obtuvieron 60 espectros procedentes de 6 alícuotas de cada submuestra de grasa de los 10 lotes estudiados ($T_{5,0}$; $T_{4,1}$; $T_{3,2}$; $T_{2,3}$; $T_{1,4}$ y $T_{0,5}$). Estos espectros procedentes de cada una de las 6 alícuotas se diferenciaban entre sí en el número de veces que el tubo en el que estaban contenidas había sufrido un proceso de congelación-descongelación.

Para la evaluación del efecto del tiempo de conservación de las muestras bajo congelación se compararon 60 espectros de 10 submuestras analizadas por primera vez a lo largo de 5 meses ($T_{0,0}$ a $T_{5,0}$). Para el estudio de la influencia del efecto de congelación-descongelación sobre la grasa fundida se compararon 60 espectros obtenidos de 10 submuestras conservadas durante un mismo periodo de congelación (5 meses) pero sometidas a un número diferente de procesos de congelación y descongelación ($T_{5,0}$; $T_{4,1}$; $T_{3,2}$; $T_{2,3}$; $T_{1,4}$ y $T_{0,5}$).

Los espectros se obtuvieron mediante un espectrofotómetro Foss NIRSystems 6500 SY I equipado con módulo de giro, empleando la modalidad de análisis de doble transmisión de acuerdo con la metodología descrita en el apartado 3.2.1.2.

3.2.3.3 Tratamiento quimiométrico de datos espectrales

Para el análisis de los datos espectrales procedentes de este estudio se emplearon dos estrategias diferentes. La primera de ellas está basada en el cálculo del estadístico $RMS(C)$ utilizando el programa CONTRAST del paquete estadístico ISI ver. 3.11. La segunda estrategia se basa en el cálculo de la distancia H de Mahalanobis, a través del programa CENTER del mismo paquete estadístico.

El programa CONTRAST se utilizó para realizar la comparación entre dos espectros cuantificando las diferencias entre los mismos para cada una de las longitudes de onda. De los diferentes estadísticos de cuantificación de las diferencias entre dos espectros que ofrece el programa CONTRAST, se han empleado los datos de la raíz cuadrada del cuadrado medio del error corregido con el sesgo ($RMS(C)$), calculado

dicho estadístico para las regiones del Visible (400 a 1.098 nm) y NIR (1.100 a 2.500 nm). El estadístico $RMS(C)$ viene definido por la siguiente fórmula (Shenk y Westerhaus, 1995b):

$$RMS(C) = 10^6 \times \sqrt{\frac{\sum_{i=1}^n (y_{ij} - y_{ik})^2 - \frac{\left(\sum_{i=1}^n (y_{ij} - y_{ik})\right)^2}{n}}{n-1}}$$

donde

y_{ij} = Valor de $\log(1/R)$ de la submuestra j para la longitud de onda i (\mathbf{I}_i).

y_{ik} = Valor de $\log(1/R)$ de la submuestra k para la longitud de onda i (\mathbf{I}_i).

n = número de longitudes de onda empleadas.

No se ha eliminado ninguna banda específica de los espectros para el cálculo de dicho estadístico, por lo cual, el número de longitudes de onda empleadas fue de 350 al emplear el rango del Visible y 700 al considerar únicamente el rango NIR. Como se puede deducir de la fórmula anterior, valores altos de $RMS(C)$ indicarán diferencias notables entre espectros; asimismo un valor cero de $RMS(C)$ implicaría que los dos espectros usados para su cálculo son idénticos. El hecho de introducir un factor multiplicativo de 10^6 en la fórmula anterior es para obtener un rango numérico de dicho estadístico más manejable.

El estadístico $RMS(C)$ se aplicó a espectros procedentes de alícuotas sometidas a diferentes periodos de tiempo bajo congelación como a espectros de alícuotas que hubieran sufrido un número distinto de tratamientos de congelación-descongelación.

El programa CENTER permite calcular la distancia H estandarizada, variante de la distancia de Mahalanobis (Mahalanobis, 1936), existente entre espectros en un espacio vectorial de n dimensiones, siendo n el número de nuevas variables (componentes principales) definidas a partir de un análisis de componentes principales de los datos espectrales (variables originales). En este nuevo espacio vectorial, cada uno de los espectros vendrá definido por un punto con diferentes coordenadas (*scores* o valores propios) sobre cada uno de los ejes definidos por las componentes principales consideradas. De acuerdo con Shenk y Westerhaus (1995b), la fórmula empleada por el programa para el cálculo de la distancia H es:

$$H = z \times (Z' \times Z)^{-1} \times z'$$

donde z representa el vector de los valores propios (*scores*) de una determinada muestra en el espacio de componentes principales mientras que Z es la matriz de valores propios de las muestras del colectivo de calibración en el mismo espacio vectorial.

Mediante el algoritmo CENTER, es posible el desarrollo, a partir de un conjunto de espectros A , de un espacio vectorial n dimensional (siendo n el número de componentes principales empleadas). A continuación se puede proyectar una población espectral B sobre el espacio vectorial definido por la población A , procediendo a continuación al cálculo la distancia H de cada uno de los espectros y del centroide de la población B al centroide de la población A . Cuanto menor sea la distancia entre ambos centroides, más parecidas serán ambas poblaciones en el espacio n dimensional.

Así, en este trabajo se calcularon, mediante esta última modalidad, diferentes estadísticos H dependiendo del rango de variables originales empleadas en el cálculo de las componentes principales. Así, se denominará $H_{VIS+NIR}$ cuando se calcule la distancia sobre todo el rango de medida (400 a 2500 nm), H_{VIS} cuando las variables originales empleadas sean las incluidas en el intervalo de 400 a 1100 nm y H_{NIR} en el caso de que se calcule la distancia H en el rango de 1100 a 2500 nm.

Al igual que para el estadístico $RMS(C)$, se procedió al cálculo del estadístico H a partir de espectros procedentes de alícuotas sometidas a diferentes periodos de tiempo bajo congelación como a espectros de alícuotas que hubieran sufrido un número distinto de tratamientos de congelación-descongelación.

3.3 DESARROLLO Y EVALUACIÓN DE CALIBRACIONES NIRS PARA LA PREDICCIÓN DEL CONTENIDO EN ÁCIDOS GRASOS

3.3.1 MODALIDAD DE DOBLE TRANSMISIÓN

3.3.1.1 Muestras

En este estudio se han empleado muestras de grasa subcutánea de cerdos Ibéricos procedentes de partidas amparadas por la Denominación de Origen “Jamón de Huelva” a lo largo de las campañas 96/97, 97/98 y 98/99.

Como colectivo de calibración se empleó un total de 341 muestras de grasa, cada una de las cuales fue extraída, mediante la metodología de fusión en horno microondas descrita en el apartado 3.1.1.2, de muestras de tejido adiposo procedentes de un mismo número de lotes de cerdos Ibéricos. El número de muestras representativas de cada lote se fijó en un porcentaje de acuerdo con los resultados del ensayo descrito en el apartado 4.1.2 y de la zona anatómica recomendada tras el estudio descrito en el apartado 4.1.1.

Como colectivos de validación se emplearon 3 grupos diferentes de muestras: A, B y C. Las principales diferencias entre ellos radicaban en las características de las muestras y su periodo de análisis NIRS, como se indicará en el apartado 3.3.1.3.

3.3.1.2 *Análisis químico de referencia*

A las muestras de grasa se les determinó el contenido de ésteres metílicos de ácidos grasos mediante cromatografía de gases en el Laboratorio Agroalimentario de la Junta de Andalucía en Córdoba, de acuerdo con la metodología descrita en el apartado 3.1.1.3. En concreto, las condiciones cromatográficas empleadas en dicha determinación fueron: uso de inyección automática de 0,5 μ l de muestra con una relación de split de 1:90, empleo de nitrógeno como gas portador a 25 ml/min, temperaturas de inyector y detector FID a 250°C, uso de temperatura programada durante el análisis (165°C durante 19 min, incremento de 5°C/min y 200°C durante 7 min) y utilización de columna BPX70 de 50 m de longitud, con 0,22 mm de diámetro interno y un espesor de película de 0,25 μ m.

3.3.1.3 *Obtención de datos espectroscópicos NIR*

Previo al envío de las muestras de grasa al laboratorio para su análisis mediante cromatografía de gases, se obtuvieron los espectros NIRS por duplicado de dichas muestras.

De acuerdo con la metodología descrita en el apartado 4.2.1, el análisis NIRS de las muestras de calibración se realizó en el periodo comprendido entre octubre de 1.997 y abril de 1.998. Los espectros NIR se obtuvieron de acuerdo con el protocolo descrito en el apartado 3.2.1, en el rango de 400 a 2.500 nm con intervalos de 2 nm, mediante un espectrofotómetro Foss NIRSystems 6500 SY I, equipado con un módulo de giro. Para la presentación de las muestras se empleó la modalidad de doble transmisión mediante el uso de una cápsula de 0,1 mm de paso óptico y fondo de material reflectante (nº catalogo IH-0345). Para la recogida de datos espectrales se empleó el software ISI NIRS 3 versión 2.05 (Infrasoft International, Port Matilda, PA, USA).

Como ya se ha indicado previamente, se emplearon 3 grupos diferentes de muestras (A, B y C) como colectivos de validación. Las principales diferencias entre ellos radicaban en las características de las muestras y su periodo de análisis NIRS.

El colectivo de validación A contenía espectros de 20 muestras de grasa de cerdo Ibérico no incluidas en el colectivo de calibración. Los espectros del colectivo A de validación fueron obtenidos en mayo de 1.998, es decir, justo después de finalizar la recogida de datos espectrales del colectivo de calibración (abril de 1.998).

El colectivo de validación B contenía espectros de 30 muestras incluidas en el colectivo de calibración pero obtenidos en noviembre de 1.999, 20 meses después de finalizar la recogida espectral del grupo de calibración.

El colectivo de validación C contenía 150 espectros obtenidos utilizando alícuotas de una única muestra de grasa de cerdo Ibérico (denominada *Grasa Control*), y que fue seleccionada por presentar una composición en ácidos grasos similar a la composición media del colectivo de calibración. Estos 150 espectros de la muestra de Grasa Control fueron obtenidos recogiendo el espectro NIR de una alícuota de dicha muestra 3 veces por semana en el periodo comprendido entre junio de 1.999 y julio de 2.000.

3.3.1.4 Tratamiento quimiométrico de datos NIRS

El tratamiento quimiométrico necesario para el desarrollo y evaluación de calibraciones se realizó utilizando el paquete quimiométrico ISI NIRS 3 ver. 3.11. La metodología utilizada es la recomendada en diferentes publicaciones (Shenk *et al.*, 1989; Martens y Næs, 1989; Mark y Workman, 1991; Moya, 1993; Shenk y Westerhaus, 1995a,b y 1996; Garrido *et al.*, 1996).

Previamente a la obtención de las ecuaciones de calibración, se aplicó el algoritmo CENTER del paquete quimiométrico ISI a los datos espectrales del colectivo de calibración. La finalidad de esta etapa fue la de reducir la dimensionalidad de dicho colectivo para, con posterioridad, poder evaluar la presencia de muestras anómalas espectrales en dicho colectivo, mediante el cálculo de la distancia H de cada muestra al centro de la población.

Este estadístico H , como fue descrito en el apartado 3.2.3.3, es una variante de la distancia de Mahalanobis y proporciona información acerca de la distancia, dentro del espacio multivariante definido por las componentes principales, entre cada muestra

(espectro) y el centroide de un grupo de muestras. La fórmula empleada para el cálculo del estadístico H es fue la descrita en el apartado 3.2.3.3 (Shenk y Westerhaus, 1995b).

Se utilizó el valor de $H=3$ como valor límite para considerar a un espectro como anómalo, es decir, aquel espectro que difiere sensiblemente del resto de la población de calibración. Aquellos espectros anómalos fueron eliminados, repitiendo el proceso de detección y eliminación de anómalos espectrales hasta lograr que todas las muestras presentaran un valor de H inferior al valor máximo recomendado.

Se evaluó el efecto sobre el cálculo del estadístico H de diferentes pretratamientos de la señal espectroscópica NIR tales como la derivación y la corrección del fenómeno de radiación dispersa (*scatter*).

Para el tratamiento de derivadas de los espectros NIR, el programa ISI utilizado en este trabajo emplea una notación de 4 dígitos (a,b,c,d) para expresar el tipo de derivadas aplicado. El primer dígito (a) se refiere al orden de magnitud de la derivada (0 = espectro sin derivar, 1 = 1ª derivada, 2 = 2ª derivada, etc.). El segundo dígito (b) indica el tamaño del segmento sobre el cual se realiza la derivación (intervalo expresado en nanómetros donde se realiza el cálculo de la derivada). Los dígitos tercero (c) y cuarto (d) muestra el tamaño de los intervalos, expresados en nanómetros, empleados para el cálculo de suavización de la derivada (Shenk y Westerhaus, 1995b).

Junto a este tratamiento matemático de derivadas, se aplicó un pretratamiento de los datos espectrales para corregir el fenómeno de radiación dispersa denominado Standard Normal Variate (*SNV*) and Detrend (*DTR*) (Barnes *et al.*, 1989 y 1994; Dhanoa *et al.*, 1994 y 1995).

El pretratamiento *SNV* centra cada espectro y lo corrige con su propia desviación típica, corrigiendo con ello desviaciones a lo largo del eje de valores de absorbancias. En numerosas referencias bibliográficas, *SNV* es también denominado como “autoescalado a varianza 1”. Su cálculo matemático se realiza de acuerdo con las fórmulas indicadas por Fernández-Cabanás y Garrido-Varo (1999):

$$Y_{SNV,i} = \frac{Y_{\log(1/R),i} - \bar{Y}_{\log(1/R)}}{S_{\log(1/R)}}$$

donde $Y_{SNV,i}$ representa el valor de absorbancia a una longitud de onda i corregido tras el tratamiento SNV, $Y_{\log(1/R),i}$ es el valor de absorbancia original a una longitud de onda i , $\bar{Y}_{\log(1/R)}$ es el valor medio de los valores de absorbancia para todas las longitudes de onda y $S_{\log(1/R)}$ representa la desviación típica de los valores de absorbancia para todas las longitudes de onda.

Por el contrario, el pretratamiento *DTR* realiza un ajuste polinomial de segundo orden entre los valores espectrales o absorbancias (variables dependientes) y las longitudes de onda (variables independientes). Con ello se corrigen las desviaciones de la línea base en el espectro. El cálculo matemático se realiza de acuerdo con la fórmula:

$$\hat{Y}_{\log(1/R),i} = A + B.W + C.W^2$$

$$Y_{DT,i} = Y_{\log(1/R),i} - \hat{Y}_{\log(1/R),i}$$

donde $Y_{DT,i}$ es el valor de absorbancia a una longitud de onda i corregido tras el tratamiento *DT*, $\hat{Y}_{\log(1/R),i}$ representa el valor estimado de $Y_{\log(1/R),i}$ obtenido mediante una función cuadrática y W es el valor de longitud de onda considerado.

El pretratamiento de corrección de la radiación difusa *SNV* y *DTR* (*SNV+DTR*) combina ambos tipos de correcciones.

Las ecuaciones de calibración fueron generadas a partir de dos métodos de regresión diferentes: regresión lineal múltiple (RLM) y una modificación de la regresión por mínimos cuadrados parciales (RMCPM) incluidos ambos tipos en el programa BCAL del paquete quimiométrico ISI ver. 3.11.

El método de regresión RLM utilizado fue el denominado “paso a paso”. Este método se basa en la elección, para cada variable, de un número máximo de longitudes de onda mediante un proceso paso a paso (*stepwise*). Estas longitudes de onda son seleccionadas mediante un test F, el cual informa de aquellas que están más correlacionadas con los valores de referencia. Una vez que una longitud de onda

concreta ha sido seleccionada e incluida en el modelo, esta es evaluada de nuevo junto al resto de longitudes de onda del espectro (Hruschka, 1987).

El proceso se repite hasta que las longitudes de onda que se añaden al modelo muestran escasa o nula mejora en la exactitud de la ecuación de calibración generadas, tratando de evitar el fenómeno de sobreajuste (Martens y Næs, 1989). Esta escasa o nula mejora en cuanto a exactitud ocurrirá cuando el valor de F estimado para una longitud de onda sea inferior al valor de F límite considerado (F límite = 8, valor recomendado por el propio programa). Se generaron ecuaciones NIRS para un máximo de 10 y 2 longitudes de onda seleccionadas por variable con la finalidad de evaluar la influencia de este factor sobre la precisión de los modelos obtenidos.

El algoritmo RMCP lleva a cabo una regresión multivariante en la que se reducen los datos espectrales a una serie de variables fundamentales (factores RMCP) que no sólo resumen la mayor parte de la información espectral sino que también son las que mayor correlación alcanzan con los valores de referencia (ácidos grasos). En la regresión RMCPM los residuos obtenidos después del cálculo de cada término de regresión son estandarizados (dividiendo por la desviación típica) antes de calcular el siguiente término de regresión.

Para calcular el número óptimo de términos de la regresión, se utilizó la validación cruzada, algoritmo que selecciona diferentes colectivos de calibración y validación dentro del total de la población considerada, realizando con cada selección una simulación del algoritmo de regresión (Shenk y Westerhaus, 1991; Martens y Næs, 1989). Cada simulación desarrolla ecuaciones con n términos, desde 1 hasta el número máximo aconsejado (generalmente uno por cada diez muestras más dos o tres términos adicionales). Finalmente el programa de cálculo selecciona la ecuación que hace mínimo el error típico de validación cruzada (*ETVC*) el cual se describirá a continuación.

Los estadísticos utilizados para la evaluación de las ecuaciones de calibración obtenidas y la selección de aquellas con mayor precisión y exactitud fueron:

- **Sesgo:** informa acerca de la media de las diferencias existentes entre el valor del constituyente determinado por el método de referencia (y_i) y el valor predicho por la ecuación de calibración (\hat{y}_i), para la muestra i . Su fórmula es:

$$sesgo = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)}{N}$$

siendo N el número de muestras del colectivo de calibración.

- **Error típico del método de referencia (ETL):** representa el error típico de muestras analizadas por duplicado mediante el método de referencia. Su fórmula, en el caso de análisis de duplicados, es:

$$ETL = \sqrt{\frac{\sum_{i=1}^M (y_1 - y_2)^2}{2M}}$$

donde y_1 e y_2 son los valores obtenidos del constituyente a calibrar para una determinada muestra y M representa el número de muestras analizadas por duplicado mediante el método de referencia.

- **Error típico de calibración (ETC):** es el error típico de los residuales para el colectivo de calibración. Su fórmula es:

$$ETC = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - p - 1}}$$

donde N representa el número de muestras del colectivo de calibración, p el número de factores o términos de la regresión, y_i el valor del constituyente a calibrar para la muestra i e \hat{y}_i el valor estimado por la ecuación de calibración para la muestra i .

- **Error típico de validación cruzada (ETVC) y error típico de predicción (ETP):** representan el error típico de los residuales para el colectivo de validación cruzada o de predicción respectivamente. Sus fórmulas son:

$$ETVC \cong ETP = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N-1}}$$

donde N representa el número de muestras empleadas como colectivo de validación cruzada o de predicción.

- **Coeficiente de determinación (R^2):** muestra el porcentaje de la variabilidad existente en las concentraciones, modeladas por la ecuación de regresión. Se emplea tanto en el proceso de calibración (R^2) como en el proceso de validación cruzada (r^2). Su fórmula es:

$$R^2 = \frac{SC_{reg}}{SC_{tot}} = \frac{SC_{tot} - SC_{res}}{SC_{tot}} = 1 - \frac{SC_{res}}{SC_{tot}} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

siendo SC_{reg} la suma de cuadrados de la regresión, SC_{res} la suma de cuadrados de los residuales, SC_{tot} la suma de cuadrados total y N el número de muestras del colectivo.

- **RPD:** relaciona la desviación típica de los datos de referencia para el colectivo de validación y el error típico de validación cruzada. Su fórmula es:

$$RPD = \frac{DT}{ETVC}$$

- **RER:** muestra la relación entre el intervalo de composición de los datos de referencia para el colectivo de calibración y el error típico de validación cruzada. Su fórmula es:

$$RER = \frac{Intervalo}{ETVC} = \frac{y_{max} - y_{min}}{ETVC}$$

siendo y_{max} e y_{min} el valor máximo y mínimo de composición, respectivamente, del colectivo de calibración.

Se emplearon dos criterios, de forma conjunta, para la eliminación de muestras anómalas durante el desarrollo de las ecuaciones de calibración NIRS. El primero de ellos evalúa la presencia de muestras anómalas en cuando a su composición

química. Para ello, se emplea el estadístico T , similar a una t de Student. Este estadístico compara, para cada muestra, el valor del constituyente obtenido mediante el método de referencia (y_i) y el valor predicho por la ecuación de calibración NIRS (\hat{y}_i) mediante la fórmula:

$$T = \frac{(y_i - \hat{y}_i)}{ETC} = \frac{Residual}{ETC}$$

Se consideraron muestras anómalas en cuanto a su composición química, aquellas que presentaron valores del estadístico T superiores a 2,5 (Mark y Workman, 1991).

El segundo criterio empleado evalúa la presencia de muestras anómalas en cuanto a su información espectral NIR. Para ello, se utilizó el estadístico H , similar al empleado tras el análisis de componentes principales previo a la calibración. Sin embargo, en este caso el estadístico H es calculado a partir de un espacio vectorial que utiliza como variables los factores RMCP de la regresión obtenida para cada ácido graso. La fórmula empleada para su cálculo es la misma expuesta en el apartado 3.2.3.3 si bien, en este caso, z representaría el vector de los valores propios de una determinada muestra en el espacio de factores RMCP, mientras que Z es la matriz de valores propios de las muestras del colectivo de muestras en el mismo espacio vectorial.

Dado que mediante el algoritmo CENTER aplicado al colectivo de calibración se realiza una etapa previa de detección y eliminación de muestras anómalas espectrales mediante un análisis de componentes principales, se recomienda el uso de un valor de H límite menos estricto que el utilizado durante el análisis de componentes principales ($H < 3$). Así, el valor de H límite utilizado en este caso fue de 10.

Este proceso de detección y eliminación de anómalos químicos y espectrales se repitió durante el desarrollo de cada ecuación de calibración hasta lograr que todas las muestras presentaran valores de T y H inferiores a los máximos recomendados.

En todas las ecuaciones de calibración generadas en este apartado se aplicó un pretratamiento $SNV+DTR$ para la corrección del fenómeno de radiación dispersa y diferentes tratamientos matemáticos de derivadas para optimizar la precisión de los modelos generados.

Teniendo en cuenta el elevado número de ecuaciones a desarrollar, para la obtención de las mismas se ha empleado una rutina o secuencia automática incluida en el programa ISI. Este procedimiento o macro permite programar automáticamente la secuencia deseada de opciones de regresión, mediante el programa denominado AUTOCAL. En este caso se programaron 10 tipos diferentes de tratamientos matemáticos de derivadas (Tabla 3.2). Esta secuencia automática se aplicó al generar ecuaciones, tanto mediante regresión RMCPM como RLM. En todos los casos se empleó el pretratamiento matemático para la corrección del fenómeno de radiación dispersa del tipo *SNV+DTR*. Como opciones de la regresión RMCPM se incluyeron el cálculo de un máximo de 15 términos de regresión o factores RMCP, el uso de 4 grupos de validación cruzada y la aplicación de un máximo de 9 etapas de eliminación de datos anómalos. En la regresión RLM se consideró un máximo de 10 y 2 longitudes de onda seleccionadas por variable y la aplicación de nuevo de un máximo de 9 etapas para la eliminación automática de datos anómalos.

Tabla 3.2. Opciones incluidas en la secuencia automática de calibración (AUTOCAL).

Ecuación	Derivada	Corrección radiación difusa
1	0,0,1,1	<i>SNV + DTR</i>
2	1,4,4,1	<i>SNV + DTR</i>
3	1,10,5,1	<i>SNV + DTR</i>
4	1,10,10,5	<i>SNV + DTR</i>
5	2,5,5,1	<i>SNV + DTR</i>
6	2,10,5,1	<i>SNV + DTR</i>
7	2,10,10,1	<i>SNV + DTR</i>
8	3,5,5,1	<i>SNV + DTR</i>
9	3,10,5,1	<i>SNV + DTR</i>
10	3,10,10,1	<i>SNV + DTR</i>

Las mejores ecuaciones de calibración NIRS fueron seleccionadas de acuerdo con los valores para los estadísticos r^2 y *ETVC*, esto es, aquellos mayores valores de r^2 y menores valores de *ETVC*. Una vez seleccionada la mejor ecuación para un determinado constituyente, la misma fue evaluada utilizando para ello un colectivo diferente al utilizado para la obtención de la calibración (validación externa).

El paquete quimiométrico ISI, utilizado en este trabajo, tiene incorporado el procedimiento de evaluación externa desarrollado por Shenk *et al.* (1989) en el

programa MONITOR. Básicamente, dicho procedimiento de validación externa se basa en los siguientes criterios:

1. La calibración evaluada ha de tener, al menos, 100 muestras.
2. El colectivo de validación ha de tener, como mínimo, 9 muestras.
3. Se establece como límite para el sesgo el obtenido de acuerdo con la siguiente fórmula:

$$\text{Límite sesgo} = ETC \times 0,6$$

4. Se define el error típico de predicción corregido por el sesgo ($ETP(C)$) como un estimador del error aleatorio o error no explicado por la regresión. Se calcula a partir de la siguiente fórmula:

$$ETP(C) = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i - sesgo)^2}{N - 1}}$$

Al mismo tiempo, se establece un límite para el $ETP(C)$ de acuerdo con la fórmula:

$$\text{Límite } ETP(C) = ETC \times 1,3$$

5. La pendiente de la regresión entre los valores predichos y los valores del método de referencia, debería estar comprendida entre 0,9 y 1,0.
6. La media de los valores predichos y de referencia no deben diferir en más de un 20%.
7. El coeficiente de determinación (R^2) de la regresión establecida entre los valores predichos y los del método de referencia debe ser mayor de 0,6 e idealmente debería aproximarse a 1,0.
8. La distancia H de las muestras del colectivo de validación debe de ser inferior al valor límite de 3.

3.3.2 MODALIDAD DE INTERACTANCIA-REFLECTANCIA

3.3.2.1 Muestras

En este estudio se han empleado muestras de tejido adiposo subcutáneo de cerdos Ibéricos sacrificados durante las campañas 97/98 a 99/00 y procedentes de la Asociación Española de Criadores de Ganado Porcino Tronco Ibérico (A.E.C.E.R.I.B.E.R.), de la empresa Sánchez Romero Carvajal – Jabugo, S.A. y de la Sdad. Coop. Ganadera del Valle de Los Pedroches (COVAP).

Como colectivo de calibración se empleó un total de 164 muestras de tejido adiposo, obtenido de la zona anatómica seleccionada en el apartado 4.1.1, de un número similar de cerdos Ibéricos.

Como colectivo de validación, denominado colectivo D, se emplearon espectros procedentes de 15 muestras de grasa de cerdo Ibérico no incluidas en el colectivo de calibración, los cuales procedían de la Asociación Española de Criadores de Ganado Porcino Tronco Ibérico (A.E.C.E.R.I.B.E.R.).

3.3.2.2 Análisis químico de referencia

Tanto en las muestras del colectivo de calibración como en el de validación se determinó el contenido de ésteres metílicos de ácidos grasos mediante cromatografía de gases en el Laboratorio Agroalimentario de la Consejería de Agricultura y Pesca de la Junta de Andalucía en Córdoba, de acuerdo con la metodología descrita en el apartado 3.1.1.3.

3.3.2.3 Obtención de datos espectroscópicos NIR

El análisis NIRS de las muestras de tejido adiposo de los colectivos de calibración y validación se realizó en el periodo comprendido entre noviembre de 1.997 y marzo de 2.000. Los espectros NIR se obtuvieron de acuerdo con el protocolo descrito en el apartado 4.2.2, mediante un espectrofotómetro Foss NIRSystems 6500 SY II equipado con un módulo de sonda de fibra óptica (modalidad de interactancia-

reflectancia). El rango espectral fue de 800 a 2.200 nm con intervalos de 2 nm. Para la recogida de datos espectrales se empleó el software ISI NIRS 3 versión 3.11.

3.3.2.4 Tratamiento quimiométrico de datos NIRS

El tratamiento quimiométrico de datos espectrales y químicos se realizó con el paquete informático ISI NIRS 3 ver. 3.11. La metodología a seguir para el trabajo de desarrollo y evaluación de calibraciones fue la misma empleada en el apartado 3.3.1.4.

Como etapa previa a la obtención de las ecuaciones de calibración, se procedió al desarrollo de un análisis de componentes principales de los datos espectrales, evaluando de nuevo la influencia sobre los resultados de la aplicación de pretratamientos matemáticos ($SNV+DTR$ y derivadas) en la señal espectral.

Las ecuaciones de calibración fueron generadas mediante regresión lineal múltiple (RLM) y una modificación de la regresión por mínimos cuadrados parciales (RMCPM) incluidos ambos tipos en el programa ISI.

Se emplearon los mismos pretratamientos matemáticos de corrección de la radiación dispersa, tratamientos de derivadas de los espectros, estadísticos de selección y evaluación de ecuaciones que los indicados en el apartado 3.3.1.4.

3.3.3 CONTROL Y DIAGNÓSTICOS DE RUTINA DE LA INSTRUMENTACIÓN NIRS UTILIZADA

Durante todo el periodo de obtención de datos espectroscópicos (de octubre de 1.997 a julio de 2.000), los instrumentos NIRS utilizados en este estudio han estado sometidos a la rutina de control y diagnósticos recomendados por el fabricante (Shenk y Westerhaus, 1995b).

Los controles empleados han consistido en el análisis diario de una muestra de referencia (denominada *Check Cell*) y la realización del grupo de pruebas que constituyen los diagnósticos denominados repetibilidad de la señal analítica, exactitud de la longitud de onda y medida de la respuesta de los detectores del instrumento.

Dichos controles se han realizado semanalmente y cuando se ha sospechado sobre la presencia de algún problema instrumental y previamente a algún trabajo de clonación de equipos.

El análisis de la muestra de referencia y la evaluación de los diagnósticos deben realizarse una vez que el instrumento se ha estabilizado, aproximadamente tras permanecer una hora encendida la lámpara del instrumento.

3.3.3.1 Repetibilidad de la señal espectroscópica NIR

La repetibilidad de la señal espectroscópica es una medida del denominado ruido fotométrico (McClure, 1987). El ruido fotométrico en un espectrofotómetro es una medida del ruido total existente en el sistema, el cual es originado por los detectores, amplificadores y el convertidor analógico/digital del equipo (McClure, 1987).

Dicho ruido se calcula obteniendo dos espectros de un material de referencia (en este caso, cerámica) una vez analizado como tal y otra como si se tratara de una muestra. Posteriormente, ambos espectros se restan y resulta lo que se denomina “espectro del ruido” ó δ_n . Una vez obtenido dicho espectro, el ruido se cuantifica de forma independiente para cada intervalo de medida (de 400 a 1.098 y de 1.100 a 2.500 nm), utilizando los estadísticos sesgo, raíz cuadrada del cuadrado medio del error (*RMS*) y raíz cuadrada del cuadrado medio del error corregido por el sesgo (*RMS(C)*), siendo:

$$sesgo = 10^6 \times \frac{\sum_{i=1}^n d_i}{n}$$

$$RMS = 10^6 \times \sqrt{\frac{\sum_{i=1}^n (d_i)^2}{n}}$$

$$RMS(C) = 10^6 \times \sqrt{\frac{\sum_{i=1}^n (d_i)^2 - \frac{\left(\sum_{i=1}^n d_i\right)^2}{n}}{n-1}}$$

siendo d_i las diferencias entre valores de absorbancia a una longitud de onda determinada y n el número de datos espectrales.

Este test se realiza mediante la opción “Measure NIR Repeatability” incluida en el menú “Diagnostics” del programa SCAN de ISI NIRS3 ver. 2.05. La secuencia de

análisis del material de referencia se repite varias veces (se recomienda entre 5 y 10 veces) y se calcula el valor medio de los diferentes estadísticos.

Considerando únicamente la región NIR en un instrumento equipado con el módulo de giro, el valor medio del estadístico $RMS(C)$ ha de ser inferior a $20 \mu\log(1/R)$, en condiciones estables de temperatura y humedad. En caso contrario, el valor medio de dicho estadístico puede ser superior. En la región NIR, el sesgo ha de ser, en valor absoluto, inferior a $10 \mu\log(1/R)$ en condiciones estables y, en ningún caso, ha de presentar valores superiores a $100 \mu\log(1/R)$ (Shenk y Westerhaus, 1995b).

Es importante indicar que el valor límite de repetibilidad para la región NIR es superior cuando se trabaja con un instrumento equipado con un módulo de sonda de fibra óptica debido a la presencia de vibraciones y pérdida de señal en la fibra óptica, entrada de luz externa, etc. En este caso, el valor límite recomendado por el fabricante es $100 \mu\log(1/R)$.

3.3.3.2 *Exactitud de la longitud de onda*

La exactitud de la longitud de onda es medida por comparación con un número determinado de picos de absorción a longitudes de onda características de un material de referencia interno (didimio para el rango del Visible y poliestireno para la región NIR). Se utilizan 8 longitudes de onda nominales diferentes. Tres de ellas, características del didimio (805,23; 878,52 y 1.143,63 nm), son empleadas para evaluar el rango del visible (400 a 1.098 nm). Las cuatro restantes, características del poliestireno (1.143,63; 1.681,27; 2.166,40 y 2.305,93 nm), son empleadas para evaluar el rango NIR (1.100 a 2.500 nm).

Este test se realiza mediante la opción “Measure Wavelength Accuracy” incluida en el menú “Diagnostics” del programa SCAN de ISI NIRS-ver. 2.05. Los parámetros utilizados para evaluar la precisión de la longitud de onda han sido los valores denominados K , f y el error.

Los valores de K y f son coeficientes matemáticos empleados en la corrección de los valores de longitud de onda aportados por el instrumento, en caso de que las longitudes de onda observadas en la media de los materiales de referencia

difieran significativamente de las teóricas. La relación empleada por la red de difracción del instrumento para la obtención de una radiación con una determinada longitud de onda es (Shenk y Westerhaus, 1995b):

$$\text{arcseno}(\mathbf{I}) = K \times \text{seno}(\mathbf{q} + \mathbf{f})$$

donde \mathbf{I} es la longitud de onda y \mathbf{q} es el ángulo de inclinación de la red de difracción.

Como puede observarse en la anterior ecuación, los valores de K y \mathbf{f} son coeficientes empleados para alinear las longitudes de onda de la radiación emitida por el instrumento a las de los picos de los materiales de referencia empleados.

Por el contrario, el valor del error informa de la diferencia media existente entre los valores de longitudes de onda teóricas o de referencia y los valores de las longitudes de onda estimadas. Por tanto, el valor del error aporta información acerca de la exactitud del instrumento. Durante la realización de este diagnóstico, el programa proporciona información de un valor de K , \mathbf{f} y el error para la situación instrumental en la que se realiza la medida (*current*) y un valor de K , \mathbf{f} y el error sugeridos (*suggested*) para mejorar la exactitud en la medida de la longitud de onda.

Se recomienda modificar los valores de K y \mathbf{f} cuando el error existente en el instrumento sea superior al doble del error sugerido, o bien cuando el error sea superior a +/-0,30 nm (Shenk y Westerhaus, 1995b).

3.3.3.3 *Respuesta de los detectores del instrumento*

La respuesta del instrumento es una medida de la reflectancia absoluta procedente de la cerámica (estándar interno para el parámetro de respuesta). El valor máximo de dicha respuesta es 65.535.

Este test se realiza mediante la opción “Measure Instrument Response” incluida en el menú “Diagnostics” del programa SCAN de ISI NIR3 ver. 2.05. Sin embargo, este test presenta unas particularidades dependiendo de las características del equipo Foss NIRSystems empleado. Así, el instrumento puede realizar un ajuste de la respuesta de los detectores de forma automática (ganancia automática) o bien es necesario realizar dicho ajuste en el equipo de forma manual. En el primer caso, el

instrumento Foss NIRSystems se denomina SY II mientras que en el segundo caso es del tipo SY I.

Los valores de respuesta recomendados por el fabricante dependen de esta particularidad. Si el instrumento dispone de detectores dotados con la opción de ganancia automática (SY II), el valor de la respuesta para el rango del visible ha de ser superior a 50.000 mientras que para el rango NIR debería ser superior a 35.000.

En caso de que los detectores no dispongan de la opción de ganancia automática (SY I), es necesario asegurarse de que los valores máximos de respuesta del instrumento se sitúen entre el 84 y 80% del máximo permitido (65.535), es decir, dentro del rango de 55.000 a 58.000 (Shenk y Westerhaus, 1995b).

Los valores de los parámetros aportados en este test dependerán por tanto del tipo de detectores. Sin embargo se ha de tener en cuenta que, para todos los casos, los valores máximos de la respuesta han de ser superiores a 30.000 (en la medida de lo posible, se fijarán en torno a 55.000) y los valores del parámetro de periodo oscuro (*dark period values*) deben ser inferiores a 1.600, siendo su valor ideal próximo a 900.

3.3.3.4 Muestra de referencia (Check Cell)

La muestra de referencia, también denominada *Check Cell*, es una muestra de un concentrado proteico molido envasado en una cápsula sellada para evitar que su composición se modifique en el tiempo. Esta cápsula es suministrada por el fabricante de los equipos Foss NIRSystems con la finalidad, entre otras, de realizar una evaluación general de la exactitud y precisión de los análisis realizados con un determinado instrumento NIRS. Para ello, se recomienda analizar esta muestra sellada antes de comenzar a obtener espectros en cada sesión de trabajo del instrumento NIRS.

Esta muestra de referencia es suministrada junto con ecuaciones de calibración desarrolladas en el equipo denominado Master ISI, para predecir valores de proteína bruta, grasa bruta, fibra bruta y materia seca de la muestra de referencia en cualquier instrumento Foss NIRSystems. Para ello es preciso previamente la realización de un proceso de estandarización, denominado CLONE1, que se lleva a cabo tras la recepción de un nuevo instrumento NIRS (Shenk y Westerhaus, 1995b).

Una vez realizada la estandarización, el instrumento NIRS puede utilizar las ecuaciones desarrolladas en el Master ISI y obtener resultados de similar precisión y exactitud que los que se obtendrían en dicho Master. Los estadísticos de las ecuaciones suministradas aparecen reflejados en la Tabla 3.3.

Tabla 3.3. Estadísticos de ecuaciones para el análisis NIRS de la muestra de referencia (Check Cell).

Constituyente	ETC	R²	ETVC	r²
Proteína Bruta (% MS)	0,75	0,96	0,81	0,95
Grasa Bruta (% MS)	0,15	0,99	0,17	0,98
Fibra Bruta (% MS)	0,48	0,83	0,58	0,76
Materia Seca (%)	0,29	0,90	0,35	0,86

El análisis de esta muestra de referencia se realizó haciendo uso de la opción “Check Cell” incluida en el programa SCAN de ISI NIRS 3 ver. 2.05. Dicha opción permite que a partir de los espectros obtenidos y haciendo uso de las ecuaciones indicadas en la Tabla 3.3, se obtenga la composición de la muestra de referencia.

Durante el periodo de desarrollo y validación de los modelos cuantitativos y cualitativos para la grasa de cerdo Ibérico (de octubre de 1.997 a julio de 2.000), esta muestra de referencia fue analizada diariamente en las mismas condiciones instrumentales y ambientales que la muestra de Grasa Control y que todas las muestras de grasa fundida de cerdo Ibérico analizadas en esta Tesis. De esta forma se obtuvieron más de 600 espectros de la muestra de referencia y las correspondientes predicciones de proteína bruta, grasa bruta, fibra bruta y materia seca.

A partir de la información química generada se obtuvo el valor medio (\bar{y}_{mr}) y la desviación típica (\mathbf{s}_{mr}) para cada constituyente de la muestra de referencia durante el periodo considerado. De acuerdo con la norma ASTM E 1655 (1994), los valores estimados para cada constituyente de una muestra de referencia han de estar dentro del rango definido por $[\bar{y}_{mr} - (t \times \mathbf{s}_{mr})]$ y $[\bar{y}_{mr} + (t \times \mathbf{s}_{mr})]$ para un nivel de probabilidad determinado, siendo t el valor de la t de Student para $(q-1)$ grados de libertad y un nivel de probabilidad concreto y q el número de espectros considerados. En el caso del programa ISI, para el cálculo de estos límites se emplean los 60 últimos valores

predichos para cada constituyente de la muestra de referencia ($q=60$) y una aproximación de 3 para el valor de la t de Student ($t=3$).

Para el estudio y discusión de los resultados obtenidos de este control instrumental se han utilizado los diagramas de control empleados frecuentemente en programas de control de calidad.

3.4 DESARROLLO Y EVALUACIÓN DE MODELOS MULTIVARIANTES DE CLASIFICACIÓN DE GRASA DE CERDO IBÉRICO

3.4.1 MATERIAL EXPERIMENTAL

3.4.1.1 Muestras

Se han empleado muestras de grasa subcutánea de cerdos Ibéricos sacrificados durante la campaña 97/98, procedentes de lotes pertenecientes a la Sdad Coop. Ganadera del Valle de Los Pedroches (COVAP).

Como consecuencia del control y seguimiento exhaustivo realizado por los Servicios Técnicos de COVAP, se pudo disponer de una amplia información de las características de los lotes experimentales y de las explotaciones ganaderas junto con una completa descripción del tipo de alimentación a que fueron sometidos los lotes de cerdos durante las etapas de precebo y cebo. Entre esta información se incluía el número de cerdos y tipo genético de cada lote, su fecha de sacrificio, peso y edad final media de los animales de cada lote, tipo de alimentación empleada durante el cebo, fecha de inicio y número de días durante los cuales se alimentaron los cerdos en montanera, cantidad de pienso total y aporte medio diario de pienso administrado a cada animal en la fase de cebo.

Para el desarrollo de los diferentes modelos multivariantes estudiados, se emplearon 5 lotes con diferente alimentación, tratando de cubrir la variabilidad existente en cuanto a tipo de alimentación durante la fase de cebo de cerdos Ibéricos.

El lote I constaba de 41 cerdos Ibéricos puros alimentados durante 6 meses en montanera, sin ningún aporte adicional de pienso, excepto durante los 10 primeros días del cebo donde se aportó una media de 1kg/día de pienso por cerdo.

El lote II lo formaban 36 cerdos Ibéricos cruzados al 75% con Duroc-Jersey, alimentados durante 4 meses en montanera recibiendo un complemento medio de pienso de 0,5 kg/día por animal durante el primer mes, y el resto (3 meses) sin aporte de pienso.

El lote III estaba constituido por 20 cerdos Ibéricos puros los cuales fueron alimentados en régimen de montanera durante 2 meses. El primer mes de dicho periodo se le administró una ración media diaria de 0,5 kg de pienso por animal, mientras que durante el resto de la montanera (1 mes) no tuvo aporte de pienso.

El lote IV estaba formado por 11 cerdos Ibéricos puros, habiéndose alimentado durante 8 meses en montanera con complemento diario elevado de pienso, en su etapa final. Así, en los 2 primeros meses del cebo, se alimentaron únicamente en montanera, continuando con un periodo de 3 meses en el que se complementó su alimentación en montanera con un aporte creciente diario de 0,5 a 4 kg de pienso y finalizando su cebo con un aporte exclusivo de pienso de 4 a 4,5 kg/día por animal.

El lote V lo constituían 13 cerdos Ibéricos cruzados al 75% con Duroc-Jersey, alimentados de forma exclusiva con pienso (*ad libitum*) en su fase de cebo.

A modo de resumen, en la Tabla 3.4 se indican las características de la alimentación durante el cebo de cada uno de los lotes empleados para el desarrollo de los modelos.

De cada uno de los cerdos de dichos lotes se obtuvieron muestras de tejido adiposo de la zona anatómica seleccionada en el apartado 4.1.1. Por tanto, se dispuso de un total de 121 muestras de tejido adiposo procedentes del mismo número de cerdos Ibéricos, para el desarrollo de los modelos multivariantes.

Tabla 3.4. Resumen del tipo de alimentación durante el cebo de los lotes empleados para el desarrollo de los modelos.

Lote	Nº cerdos	Alimentación durante cebo
I	41	10 días (1 kg/día); resto (6 meses) sin pienso
II	36	1 mes (0,5 kg/día); resto (3 meses) sin pienso
III	20	1 mes (0,5 kg/día); resto (1 mes) sin pienso
IV	11	2 meses (bellota); 3 meses (0,5 kg/día incrementándose); 3 meses (4 a 4,5 kg/día).
V	13	Pienso <i>ad libitum</i>

Junto a dicho colectivo de aprendizaje consideraron 3 lotes de cerdos (A, B y C) para la evaluación de los modelos generados. Dichos lotes constituyeron el colectivo de validación o de evaluación.

Así, el lote A estaba formado por 27 cerdos Ibéricos puros alimentados durante 4,5 meses en montanera sin ningún aporte adicional de pienso.

El lote B estaba constituido por 45 cerdos Ibéricos puros los cuales fueron alimentados en régimen de montanera durante 3,5 meses. El primer mes de dicho periodo se le administró una ración media diaria de 1 kg de pienso por animal, mientras que durante el resto de la montanera (2,5 meses) no tuvo aporte de pienso.

El lote C lo formaban 60 cerdos Ibéricos cruzados al 75% con Duroc-Jersey, alimentados durante 4 meses en montanera recibiendo un complemento medio de pienso de 1 kg/día por animal durante el primer mes, 2 meses sin ningún aporte de pienso, completando de nuevo con pienso (1 kg/día por animal) su estancia en montanera durante el último mes.

La Tabla 3.5 resume las características de la alimentación durante el cebo de cada uno de los lotes empleados para la validación de los modelos generados.

Tabla 3.5. Resumen del tipo de alimentación durante el cebo de los lotes empleados para validación.

Lote	Nº cerdos	Alimentación durante cebo
A	27	4,5 meses en montanera (sin pienso)
B	45	1 mes (1 kg/día); resto (2,5 meses) sin pienso
C	60	1 mes (1 kg/día), 2 meses sin pienso, 1 mes (1 kg/día)

Al igual que para el colectivo de calibración, de cada uno de los cerdos de dichos lotes se obtuvieron muestras de tejido adiposo de la zona anatómica seleccionada en el apartado 4.1.1, disponiendo con ello de un total de 132 muestras de tejido adiposo para la evaluación de los modelos.

De cada una de las muestras de tejido adiposo fueron extraídas, de forma individual y mediante la metodología de fusión en horno microondas descrita en el apartado 3.1.1.2, un mismo número de muestras de grasa fundida. Así, se disponía por tanto de 121 muestras de grasa fundida, procedentes de los lotes descritos en la Tabla 3.4, con las que se desarrollaron modelos multivariantes.

De forma paralela, se realizó la extracción de una muestra media representativa de cada lote, mediante fusión en horno microondas, a partir de todas las muestras de tejido adiposo procedentes de los cerdos Ibéricos de un mismo lote. Con ello se obtuvieron 5 muestras de grasa fundida, cada una de ellas media representativa de los lotes de cerdos descritos en la Tabla 3.4. A cada una de estas 5 muestras de grasa fundida se les determinó el contenido de ésteres metílicos de ácidos grasos mediante cromatografía de gases, en el Laboratorio Agroalimentario de la Consejería de Agricultura y Pesca de la Junta de Andalucía en Córdoba, de acuerdo con la metodología descrita en el apartado 3.1.1.3.

3.4.1.2 Obtención de datos espectroscópicos NIRS

De las muestras de grasa fundida y tejido adiposo descritas en el apartado anterior, se obtuvieron los espectros NIRS por duplicado de acuerdo con las metodologías expuestas en los apartados 4.2.1 y 4.2.2 respectivamente.

Los análisis NIRS de las muestras de grasa fundida se realizaron en el periodo comprendido entre enero y septiembre de 1998, mediante un espectrofotómetro Foss NIRSystems 6500 SY I equipado con un módulo de giro y en el rango de 400 a 2.500 nm con intervalos de 2 nm. Para la presentación de las muestras se empleó de nuevo la modalidad de doble transmisión, mediante el uso de una cápsula de 0,1 mm de paso óptico y fondo de material reflectante.

El análisis NIRS de las muestras de tejido adiposo se realizó en el periodo comprendido entre enero y mayo de 1.998. Los espectros NIR se obtuvieron mediante

un espectrofotómetro Foss NIRSystems 6500 SY II equipado con un módulo de sonda de fibra óptica (modalidad de interactancia-reflectancia), y en el rango de 800 a 2200 nm, con intervalos de 2 nm.

Para la recogida de datos espectrales en ambas formas de presentación de muestras se empleó el software ISI ver. 2.05.

3.4.1.3 Tratamiento quimiométrico de datos espectrales previo al desarrollo de modelos multivariantes

Como etapa previa a la obtención de modelos cualitativos multivariantes, se desarrolló un análisis de componentes principales de los datos espectrales, tanto de grasa fundida como de tejido adiposo, mediante el programa CENTER del paquete estadístico ISI ver. 3.11. La finalidad de esta etapa fue el evaluar la presencia de muestras anómalas espectrales en el colectivo de calibración, mediante el uso del estadístico H de cada muestra al centro de la población. El valor de H límite empleado para considerar a una muestra como anómala fue de 3, valor de H límite recomendado por el propio programa (Shenk y Westerhaus, 1995b).

Con la finalidad de poder comparar los resultados de clasificación obtenidos mediante modelos multivariantes, desarrollados a partir la composición en ácidos grasos predichos por NIRS y datos espectrales NIRS *per se*, se obtuvieron las predicciones de ácidos grasos del colectivo de calibración. De este modo, se desarrollaron modelos multivariantes a partir de datos de composición en ácidos grasos y datos espectrales NIRS predichos tanto en grasa fundida como en tejido adiposo.

De acuerdo con la bibliografía (Sato, 1994; Bewig *et al.*, 1994 y 1995) se recomienda, al desarrollar modelos cualitativos y cualitativos, el emplear algún tipo de pretratamiento de derivadas sobre los espectros que permita eliminar la influencia de factores físicos. Por ello, se ha aplicado un pretratamiento de derivadas del tipo (2,5,5,1) tanto sobre los datos espectrales de grasa fundida como de tejido adiposo.

Este tratamiento quimiométrico previo de datos espectrales se realizó mediante el paquete informático ISI NIRS 3 ver. 3.11.

3.4.2 ALGORITMOS DE CLASIFICACIÓN

Para el desarrollo de modelos de clasificación se han empleado diversos algoritmos de análisis multivariante cualitativo, empleando como variables tanto los valores de composición de ácidos grasos predichos como los datos espectrales NIR “per se”. La metodología a seguir para el trabajo de desarrollo y evaluación de modelos multivariantes de clasificación aparece descrita en detalle en diferentes publicaciones (Massart *et al.*, 1988 y 1997; Martens y Næs, 1989; Martens y Martens, 2000; Næs *et al.*, 2002).

Los algoritmos de análisis cualitativo empleados en este trabajo fueron el análisis discriminante canónico, lineal y cuadrático, el análisis SIMCA y el análisis discriminante basado en regresión por mínimos cuadrados parciales (RMCP).

3.4.2.1 *Análisis discriminante*

El objetivo del análisis discriminante es comprobar en qué medida es posible diferenciar correctamente las clases establecidas a partir de las variables analizadas, estableciéndose criterios estadísticos que permitan estimar la distancia entre las clases y, finalmente, asignar objetos a la clase correspondiente.

El cálculo a realizar en dicho análisis consiste en determinar una función discriminante, combinación lineal de las variables medidas, que minimizando la variabilidad existente entre las muestras de una misma clase, maximice las diferencias existentes entre las diferentes clases consideradas (Massart *et al.*, 1988 y 1997; Næs *et al.*, 2002).

Así, para una muestra desconocida, se procederá al cálculo de una distancia multivariante (normalmente, la distancia de Mahalanobis) de cada una de las muestras al centroide de las diferentes clases, a partir de las variables medidas. Cada muestra será clasificada en aquel grupo al cual presente menor distancia. Además, a partir del cálculo de las diferentes distancias, es posible determinar la probabilidad de pertenencia de una muestra a un grupo determinado (Massart *et al.*, 1988 y 1997; Næs *et al.*, 2002).

Para el desarrollo de modelos discriminantes a partir de datos espectrales y de composición en ácidos grasos, obtenida por ambas modalidades de análisis NIRS, se empleó el paquete estadístico SAS ver. 6.12.

En el desarrollo de modelos de clasificación a partir de datos espectrales NIR se emplea un elevado número de variables (valores de absorbancia) por muestra, las cuales, dadas las características de la información obtenida en la región NIR, presentan una alta correlación entre ellas. Ello provoca una información redundante en los diferentes valores de absorbancia a diferentes longitudes de onda empleados como variables (colinealidad). Una estrategia para reducir esta colinealidad consiste en la realización de un análisis quimiométrico previo de los datos mediante un análisis de componentes principales. En este tipo de análisis, se generan unas nuevas variables, denominadas componentes principales, las cuales son combinación lineal de las variables originales. Los ejes que definen estas nuevas variables o componentes principales son ortogonales, es decir, son linealmente independientes, no estando por tanto correlacionados los valores de componentes principales.

Este alto grado de correlación también se observa entre los valores predichos de ácidos grasos. Así, también fue necesario eliminar la información redundante existente en los datos de composición de ácidos grasos mediante un análisis de componentes principales. Este tratamiento previo de la información se realizó con los valores predichos de ácidos grasos a partir de las muestras de tejido adiposo y con los de grasa fundida.

Para la realización del Análisis de Componentes Principales se empleó la rutina PROC PRINCOMP del paquete estadístico SAS. (SAS, 1990), considerando la matriz de covarianzas para el cálculo de un número máximo de 15 componentes principales.

Dentro de los diferentes algoritmos de análisis discriminante que incluye el paquete estadístico SAS (SAS, 1990), se han comparado el análisis discriminante canónico, discriminante lineal y discriminante cuadrático.

El análisis discriminante canónico es empleado, además de cómo herramienta de discriminación propiamente dicha, como técnica de reducción de la dimensionalidad, similar a un análisis de componentes principales. En el análisis

discriminante canónico se generan variables canónicas que sintetizan la variación existente entre clases de la misma manera que el análisis de componentes principales sintetiza la variación total en un conjunto de datos. Así, mediante dicho análisis, se calculan unas nuevas variables canónicas, combinación lineal de las variables originales (en este caso, las componentes principales de los datos espectrales o de los valores predichos de ácidos grasos), elegidas de tal manera que se minimice la variabilidad de las muestras de una misma categoría y se maximice la variabilidad entre grupos.

El objetivo del análisis discriminante canónico es comprobar en qué medida es posible diferenciar correctamente las clases establecidas a partir de las variables analizadas. Por ello, uno de los resultados más prácticos de este tipo de análisis discriminante consiste en la representación de las muestras analizadas en un espacio definido por las nuevas variables canónicas, opción elegida en el presente trabajo para mostrar los resultados obtenidos.

En el análisis discriminante canónico no es posible asignar muestras a la clase correspondiente, opción que metodológicamente es abarcable mediante un análisis discriminante lineal o cuadrático.

Así, a partir de las variables medidas en las muestras y una variable que define las diferentes clases existentes, en el análisis discriminante lineal o cuadrático se desarrolla una función discriminante, que permita la discriminación de las muestras en las diferentes clases y la clasificación posterior de muestras ajenas al desarrollo del modelo. Esta la función discriminante, que se establece como delimitadora entre las diferentes clases, puede obtenerse o bien a partir de una única matriz de covarianzas media ponderada de todas las clases, obteniéndose con ello una función lineal, o bien a partir de las matrices de covarianzas de las diferentes clases, dando lugar a una función cuadrática. En el primer caso, el algoritmo se denomina análisis discriminante lineal y en el segundo, análisis discriminante cuadrático. El poder emplear un tipo u otro de matriz de covarianza, y por tanto de análisis discriminante, depende de la igualdad o no de las matrices de covarianzas.

Un paso previo al análisis discriminante decisivo a la hora de establecer una función lineal o cuadrática consiste en la realización de un test de homogeneidad entre las matrices de covarianzas de cada una de las clases (Morrison, 1976; SAS, 1990).

Mediante este test se estima la similitud de las matrices de covarianzas entre las diferentes clases, recomendando con ello qué tipo de análisis discriminante se debe emplear: lineal o cuadrático.

En esencia, este test de homogeneidad considera que, bajo la hipótesis nula, la expresión:

$$-2 \times RHO \times \ln \left[\frac{N^{PN/2} \times V}{\prod_{i=1}^K N(i)^{PN(i)/2}} \right] \approx \mathbf{c}^2$$

se distribuye de acuerdo con una \mathbf{c}^2 con *gl* grados de libertad, siendo

K = número de clases

P = número de variables

N = número total de observaciones – número de clases

$N(i)$ = (número de observaciones de la clase i) – 1.

$$V = \frac{\prod_{i=1}^K |\text{matriz covarianzas clase } i|^{N(i)/2}}{|\text{matriz covarianzas ponderada}|^{N/2}}$$

$$RHO = 1 - \left\{ \left[\sum_{i=1}^K \frac{1}{N(i)} - \frac{1}{N} \right] \times \frac{(2P^2 + 3P - 1)}{6(P+1)(K-1)} \right\}$$

$$gl = \frac{(K-1)P(P+1)}{2}$$

Si en el test se obtiene un valor de \mathbf{c}^2 no significativo, implica que no existen diferencias significativas al nivel de significación establecido, ante lo cual se ha de emplear la matriz de covarianzas media ponderada para el cálculo de la función discriminante lineal. Si en el test se obtiene un valor de \mathbf{c}^2 significativo para un determinado nivel, se han de emplear las matrices de covarianza individuales de cada clase, obteniéndose con ello una función discriminante cuadrática. (Morrison, 1976).

Tras llevar a cabo el test de homogeneidad de las matrices de covarianzas, se realizó el análisis discriminante lineal y cuadrático propiamente dicho.

El criterio para asignar una muestra a un grupo se expresa como la distancia de Mahalanobis de una muestra a cada uno de los grupos, por tanto, la muestra pertenecerá a aquel grupo al cual el valor de su distancia sea menor. Los estadísticos considerados tras la realización del análisis discriminante lineal y cuadrático han sido, la distancia de Mahalanobis de una muestra a cada clase y la probabilidad, en función de esta distancia, de pertenecer a una clase concreta.

Así, cuando la función discriminante es lineal (análisis discriminante lineal), la distancia de Mahalanobis se calcula mediante una matriz de covarianzas ponderada, resultante de hacer la media ponderada de cada una de las matrices de covarianzas de las diferentes clases. La fórmula matricial de la distancia de Mahalanobis (D) de una muestra x a la clase a vendría dada por la siguiente fórmula:

$$D^2_a(x) = (x - \bar{x}_a)' \times S^{-1} \times (x - \bar{x}_a)$$

donde x es el vector que define la posición de una muestra x , \bar{x}_a es el vector que define el centroide de la clase a y S es la matriz de covarianzas media ponderada.

En caso de ser una función discriminante cuadrática (análisis discriminante cuadrático), para el cálculo de la distancia de Mahalanobis se hace uso de las matrices de covarianzas de las respectivas clases. Así, la fórmula para el cálculo de la distancia de Mahalanobis de una muestra x a la clase a sería:

$$D^2_a(x) = \left[(x - \bar{x}_a)' \times S_a^{-1} \times (x - \bar{x}_a) \right] + \ln |S_a|$$

donde x es el vector que define la posición de una muestra x , \bar{x}_a es el vector que define el centroide de la clase a y S_a es la matriz de covarianzas de la clase A .

A partir de estas distancias, es posible definir la probabilidad, en %, de pertenecer una muestra x a la clase a ($\Pr(x|a)$) de acuerdo con la siguiente fórmula:

$$\Pr(x|a) = \frac{e^{-\frac{1}{2}(D_a^2(x))}}{\sum_{i=1}^K e^{-\frac{1}{2}(D_i^2(x))}} \times 100$$

donde K es el número de clases consideradas y $D_i^2(x)$ es la distancia de Mahalanobis de la muestra x a la clase i .

Para todas las posibles alternativas existentes en el presente trabajo dentro de la variabilidad de presentación de muestras (grasa fundida y tejido adiposo), se realizó un análisis discriminante lineal y cuadrático.

Para la obtención de los modelos mediante análisis discriminante canónico, lineal y cuadrático se ha utilizado la rutina PROC DISCRIM del paquete estadístico SAS ver. 6.12 (SAS, 1990). Como etapa previa al desarrollo de los modelos, ha sido necesaria la transformación de los datos espectrales obtenidos con el software ISI a un formato ASCII para su exportación al paquete estadístico SAS. Esta transformación se realizó mediante el programa WinISI ver. 1.04.

Para la evaluación de los modelos de clasificación generados mediante análisis discriminante lineal y cuadrático se ha empleado la validación cruzada completa. En este algoritmo, ya descrito en el apartado 3.3.1.4, cada una de las i muestras es clasificada mediante un modelo generado a partir de las $n-i$ muestras restantes. En el caso específico de la validación cruzada completa, se emplea una única muestra para la validación de los modelos ($i=1$), repitiéndose tantas etapas de validación cruzada como muestras posea el colectivo de clasificación.

El estadístico empleado para la evaluación de los modelos de discriminación obtenidos ha sido el error de clasificación o porcentaje de muestras clasificadas incorrectamente (en una clase diferente a su clase de origen) frente al número total de muestras del colectivo de aprendizaje. Se consideró aquel número óptimo de componentes principales mediante el cual se obtuviera un valor mínimo de error de clasificación estimado mediante validación cruzada completa, tanto para el análisis discriminante lineal como cuadrático.

3.4.2.2 Análisis SIMCA

El análisis SIMCA (Soft Independent Modelling of Class Analogy) difiere con el análisis discriminante descrito previamente en que cada clase es modelada por separado, asignándose las muestras en función de sus distancias a cada modelo. Así, en el análisis SIMCA se modela independiente cada una de las clases mediante un análisis de componentes principales, determinando con ello el centroide y la dispersión de los

valores para cada clase. La dimensión del modelo individual viene dada por el número de componentes principales que se hayan considerado.

El modelo independiente obtenido para cada clase es definido en base a dos distancias, las cuales se denominan distancia de las muestras al centro del modelo y distancia de la muestra al modelo. La primera aporta la información que queda explicada por el modelo para una determinada muestra cuando esta es proyectada sobre las componentes principales, distribuyéndose de acuerdo con una χ^2 . La segunda distancia informa sobre aquella información que no puede ser explicada por el modelo (varianza residual), representando la distancia de la muestra al modelo definido por las componentes principales. La distribución de esta distancia se ajusta a la de una F de Snedecor.

Para el desarrollo de modelos mediante análisis SIMCA se ha empleado el programa quimiométrico UNSCRAMBLER ver. 5.5, empleando como variables los datos espectrales originales (sin pretratamiento para la corrección del fenómeno de radiación difusa ni derivadas) y los valores predichos de ácidos grasos, tanto de grasa fundida como de tejido adiposo subcutáneo.

En este programa, la distancia de la muestra al centro del modelo se denomina estadístico *Leverage*, mientras que la distancia de la muestra al modelo se denomina estadístico *S*. Las condiciones para que una muestra sea clasificada como perteneciente a una determinada clase son, que tanto su distancia al centro del modelo (*leverage*) como su distancia al modelo (*S*) sean inferiores a los valores máximos definidos para ambos estadísticos en cada clase (Martens y Næs, 1989; Esbensen *et al.*, 1994).

El valor máximo del estadístico *leverage* que puede presentar una muestra para pertenecer a la clase *C* viene definido por la fórmula:

$$leverage_{max}(C) = 2 \times \frac{(a_c + 1)}{I_c}$$

donde a_c es el número de componentes principales empleadas e I_c es el número de muestras presentes en la clase *C*.

De igual modo, el valor máximo del estadístico S que puede presentar una muestra para pertenecer a la clase C a un nivel de significación, viene definido por la fórmula:

$$S_{max}(C) = S_o(C) \times \sqrt{F(k - a_c, k \times (I_c - 1))}$$

donde $S_o(C)$ es la varianza total del modelo m y F es el estadístico F de Snedecor para un nivel de significación con $(k - a_c)$ grados de libertad del numerador y $(k \times (I_c - 1))$ grados de libertad del denominador, siendo k el número de datos espectrales medidos, a_c el número de componentes principales empleadas e I_c el número de muestras de la clase C .

En la interpretación de los resultados del análisis SIMCA se ha de tener en cuenta que, dado que los modelos desarrollados para cada una de las clases son independientes, una muestra determinada puede ser clasificada en una clase, varias de ellas o bien no pertenecer a ninguna de las clases. En este último caso, las muestras que no sean clasificadas en ninguna clase, podrían dar lugar a una nueva clase si se confirman sus diferencias con el resto de los modelos.

Para la realización de este trabajo se han generado modelos para las diferentes formas de presentación de muestras consideradas y empleando ambas distancias, determinando de esta forma los modelos óptimos para cada clase y realizando con cada una la validación del modelo con el resto de toda la población muestral.

Para la importación de los datos espectrales NIRS al programa Unscrambler, se hizo uso del programa FOSS facilitado por la compañía Foss. Mediante este programa, es posible una transformación directa de los datos espectrales desde el formato empleado en los instrumentos NIRS de Foss NIRSystems a un formato ASCII específico de Unscrambler para importar datos de otros programas. Los valores predichos de ácidos grasos de grasa fundida y tejido adiposo fueron importados directamente mediante Unscrambler a partir de un formato ASCII.

Para la evaluación de los modelos de clasificación generados se procedió a la clasificación de las muestras del conjunto de clases del colectivo de aprendizaje mediante los diferentes modelos independientes de cada clase. En realidad, esta etapa no

consiste en una reclasificación ya que en la generación de cada uno de los modelos únicamente han intervenido las muestras de su clase. Dado que el análisis SIMCA es un análisis de modelado, en el que una misma muestra puede estar clasificada en una única clase, en varias o en ninguna de las clases, los estadísticos empleados han de ser diferentes al considerado en un análisis de clasificación propiamente dicho (error de clasificación). Así, los estadísticos empleados en la interpretación de resultados de un análisis SIMCA son:

- Falsos negativos: número de muestras de una determinada clase que no son clasificadas como tales por el propio modelo que define su clase.
- Falsos positivos: número de muestras que aun no perteneciendo a una clase, son clasificadas como tales por un determinado modelo.

Sin embargo, todo método de modelado puede ser empleado como método de clasificación. Para ello, basta con que cada muestra sea clasificada en aquella clase a la que presente una distancia menor. En el caso del método SIMCA, el procedimiento consistiría en clasificar las muestras en una única clase en función de sus distancias *leverage* y *S*. Así, si una muestra pertenece a una clase, se le clasifica como tal. Sin embargo, si una muestra pertenece a dos clases o bien no es asignada a ninguna clase, se clasificaría en aquella clase de la que presente menores valores de distancias *leverage* y *S*.

Al emplear el análisis de modelado como análisis de clasificación, es posible definir un error de clasificación, tanto para cada una de las clases (porcentaje de muestra mal clasificadas de una clases respecto al total de muestras de una dicha clase) como un error de clasificación global (porcentaje total de muestras mal clasificadas respecto al total de muestras empleadas en el desarrollo de modelos).

3.4.2.3 Análisis discriminante basado en regresión RMCP

El análisis discriminante basado en una regresión RMCP, a diferencia del análisis discriminante descrito en el apartado 3.4.2.1, tiene como fundamento la generación de ecuaciones NIRS discriminantes mediante un algoritmo de regresión por mínimos cuadrados parciales ó RMCP. Es decir, emplea un algoritmo comúnmente

empleado en análisis cuantitativo multivariante para resolver problemas cualitativos propiamente dichos. Para el desarrollo de modelos discriminantes basados en regresión RMCP se ha empleado el paquete quimiométrico WinISI ver. 1.04.

Como ya fue descrito en el apartado 3.3.1.4, mediante el método de regresión RMCP se generan unas nuevas variables latentes (en este caso, variables discriminatorias) que son seleccionadas no solo por ser las que mayor porcentaje de variación espectral explican sino también el mayor porcentaje de las variables de las distintas clases (Martens y Næs, 1989). De esta forma se trata, no sólo de resumir la mayor parte de información espectral, sino además relacionarlos con los valores de referencia.

Mediante este algoritmo se generan tantas variables discriminatorias como clases existan en el grupo de aprendizaje. En el caso del programa WinISI, las variables discriminatorias presentan valores entre 1 ó 2 según pertenezca una muestra a la clase o no, respectivamente. Dados estos valores, se ha considerado un valor de 1,5 como límite para establecer la pertenencia o no de una muestra a una determinada clase (Downey, 2000). Así, si una muestra presenta un valor inferior a 1,5 para una variable discriminatoria, no pertenecería a dicha clase. Por el contrario, si el valor de la variable discriminatoria es igual o superior a 1,5, la muestra sería clasificada dentro de la clase correspondiente.

Se ha de resaltar que mediante este tipo de análisis cualitativo, una misma muestra puede estar clasificada en más de una clase, al presentar valores superiores a 1,5 en varias variables discriminatorias al mismo tiempo. Esto es debido a que cada clase genera una variable discriminatoria independiente, por lo que este tipo de análisis se asemeja más a un análisis de modelado (p.ej.: SIMCA) que a un análisis discriminante propiamente dicho.

El resultado obtenido tras el desarrollo de este tipo de análisis cualitativo es una ecuación RMCP que permite estimar el grupo al cual pertenecen muestras que no han sido incluidas en el desarrollo del modelo (Shenk *et al.*, 1995a). En análisis cualitativo se nos ofrece la posibilidad de elegir entre las modalidades de regresión RMCP1 y RMCP2. La diferencia entre ambos algoritmos radica en el número de variables que se intentan discriminar. Cuando el modelo va a ser empleado para

diferenciar dos clases, el algoritmo usado es el RMCP1. En cambio cuando la discriminación se realiza con más de dos grupos el algoritmo usado debe ser el RMCP2 (Martens y Næs, 1989; Næs *et al.*, 2002). Cuando se lleva a cabo el algoritmo RMCP1 se referiría a una regresión sobre un único constituyente (la clase), simulando un análisis cuantitativo clásico y variando los valores que toma entre 1 y 2, según pertenezca o no a la clase. Sin embargo, si el algoritmo utilizado es el RMCP2 la regresión tendría lugar sobre varios constituyentes simultáneamente definiéndose en este caso una matriz cuyas columnas estarían formadas por tantas variables discriminatorias como clases haya y sus valores serían igualmente 1 y 2. En este trabajo se ha empleado un algoritmo de regresión RMCP2, incluido en el programa WinISI ver. 1.04.

Para calcular el número óptimo de términos o factores de la regresión, se ha utilizado la validación cruzada, procedimiento ya descrito en el apartado 3.3.1.4, mediante el cual se selecciona diferentes colectivos de modelado y validación dentro del total del colectivo de aprendizaje, realizando con cada selección una simulación del algoritmo de regresión discriminante. Cada simulación desarrolla ecuaciones con un número máximo de términos recomendados (generalmente uno por cada diez muestras más dos o tres términos adicionales). Finalmente el programa de cálculo selecciona la ecuación que hace mínimo el error estándar de validación cruzada (*ETVC*). Como opciones de la regresión, tanto para las ecuaciones discriminantes generadas a partir de datos espectrales o valores predichos de ácidos grasos, en grasa fundida y tejido adiposo, se indicó un número máximo de 12 términos de regresión o factores y el uso de 10 grupos de validación cruzada.

En aquellos modelos multivariantes desarrollados a partir de datos espectrales NIRS, se evaluó la influencia en el error de clasificación de la aplicación de pretratamientos de corrección de la radiación difusa (*SNV+DTR*) y diferentes tratamientos matemáticos de derivadas, ya descritos en el apartado 3.3.1.4.

Dado que el análisis discriminante basado en regresión RMCP es un análisis de modelado, similar al análisis SIMCA, una misma muestra podría estar clasificada en una, en varias o en ninguna de las clases. Por ello, los estadísticos empleados han de ser los ya mencionados en el análisis SIMCA como son el número de falsos negativos y el de falsos positivos.

Sin embargo, de igual forma que en el análisis SIMCA, el análisis discriminante basado en regresión RMCP puede ser empleado como método de clasificación. Para ello, basta con que cada muestra sea clasificada en aquella clase que presente un valor de variable discriminatoria mayor. Así, si una muestra pertenece a una única clase, se le clasifica como tal. Sin embargo, si una muestra puede pertenecer a dos clases o bien no es asignada a ninguna clase, se clasificaría en aquella clase en la que presente mayor valor de variable discriminatoria.

Al emplear el análisis de modelado como análisis de clasificación, es posible definir un error de clasificación, tanto para cada una de las clases (porcentaje de muestra mal clasificadas de una clases respecto al total de muestras de una dicha clase) como un error de clasificación global (porcentaje total de muestras mal clasificadas respecto al total de muestras empleadas en el desarrollo de modelos).

3.5 CLONACIÓN / ESTANDARIZACIÓN DE INSTRUMENTOS Y TRANSFERENCIA DE CALIBRACIONES

En este apartado se ha realizado la estandarización de varios instrumentos NIRS del mismo fabricante. Un primer monocromador Foss NIRSystems 6500, seleccionado como Master, en el cual se desarrollaron las ecuaciones para la predicción de ácidos grasos así como otros tres monocromadores Foss NIRSystems 6500, elegidos como Satélites, equipado con módulo de giro y transporte (modo estático al no producirse el giro de la cápsula) y auto ganancia de los detectores.

El equipo seleccionado como Master se encuentra ubicado en el laboratorio de la Unidad NIR/MIR del Servicio Central de Apoyo a la Investigación de la Universidad de Córdoba. En dicho laboratorio y junto al equipo anterior está localizado uno de los instrumentos Satélites.

Los dos instrumentos Satélites restantes se encuentran localizados en Zaragoza (laboratorio del Departamento de Producción Animal del Servicio de Investigación Agraria de la Diputación General de Aragón) y Cabrils (Laboratorio Agrario de la Consejería de Agricultura, Ganadería y Pesca de la Generalidad de Cataluña).

Dada esta distribución geográfica de los instrumentos, se evaluó en primer lugar la estandarización de instrumentos NIRS ubicados en Córdoba dentro del mismo laboratorio. Posteriormente, se abordó la clonación entre instrumentos localizados en laboratorios diferentes (Zaragoza y Cabrils).

3.5.1 ECUACIONES DE CALIBRACIÓN NIRS TRANSFERIDAS

Las ecuaciones de calibración NIRS empleadas en este apartado fueron similares a aquellas que se expondrán en el apartado 4.3.1.6 para la determinación del porcentaje de ácidos grasos mayoritarios en grasa de cerdo Ibérico.

Para el desarrollo de las ecuaciones NIRS se empleó el colectivo de muestras de calibración ya descrito en el apartado 3.3.1.1., analizándose las muestras de grasa líquida de dicho colectivo mediante cromatografía de gases y NIRS de acuerdo con los protocolos detallados en dicho apartado.

A modo de recordatorio, indicar que los espectros NIR se obtuvieron en el rango entre 400 y 2.500 nm, empleando para ello un monocromador Foss-NIRSystems 6500 equipado con un módulo de giro y la modalidad de análisis de doble transmisión a partir de una cápsula con fondo de material reflectante (aluminio) de 0,1 mm de paso óptico (ref. IH-0345). La composición en ácidos grasos del colectivo de calibración se describe en el apartado 4.3.1.1 (Tabla 4.3.1) si bien, para facilitar la lectura de esta Tesis se indica en la Tabla 3.6.

Tabla 3.6. Media, desviación típica e intervalo de los ácidos grasos a considerar en el colectivo de calibración (N=341).

Variable	Media	DT	Intervalo
C16:0 (%)	21,08	1,46	17,90-25,30
C18:0 (%)	10,67	1,32	7,70-14,90
C18:1 (%)	52,32	2,44	45,00-58,10
C18:2 (%)	9,39	1,32	6,80-13,50

En la obtención de las ecuaciones de calibración se empleó la metodología ya descrita en el apartado 3.3.1.4, haciendo uso del paquete informático ISI NIRS 3 ver. 3.11 para el tratamiento quimiométrico de datos espectrales y químicos.

3.5.2 MUESTRAS EMPLEADAS EN LA ESTANDARIZACIÓN

Para poder llevar a cabo la transferencia de las ecuaciones de calibración de ácidos grasos generadas en un instrumento NIRS a otro, se han empleado varios sets o

colectivos de muestras no selladas. Las características de cada uno de dichos colectivos de muestras fueron:

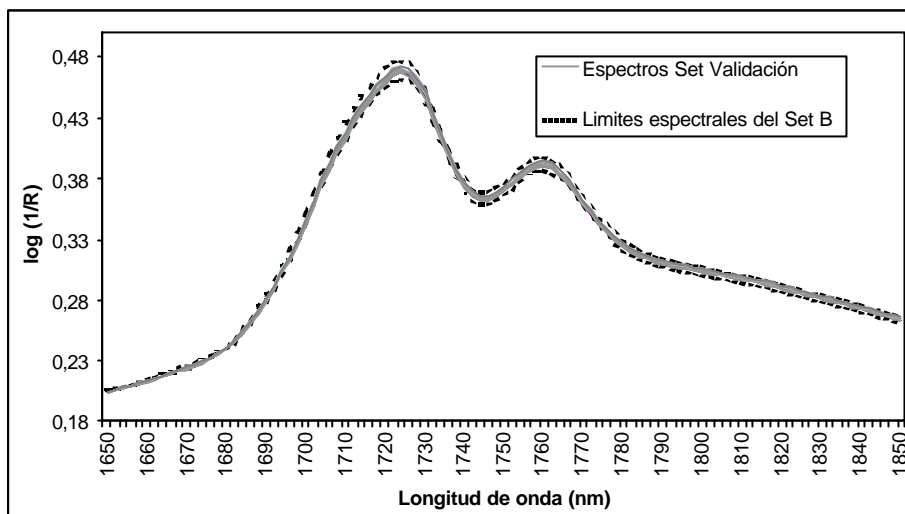
- Colectivo de Estandarización A: 1 muestra de grasa líquida de cerdo Ibérico con una composición en ácidos grasos similar a la media de la población de calibración (Tablas 3.6 y 3.7).
- Colectivo de Estandarización B: 20 muestras de grasa líquida de cerdo Ibérico seleccionadas de forma que cubrían el rango variabilidad en cuanto a densidades ópticas (valores de absorbancia) y composición en ácidos grasos del colectivo de calibración empleado para la generación de las ecuaciones (Tablas 3.6 y 3.7).
- Colectivo de Estandarización C: 1 muestra de ácido oleico (C18:1) al 95% de pureza (E. Merck, Darmstadt, Germany).
- Colectivo de Estandarización D: 1 muestra de ácido linoleico (C18:2) al 98% de pureza (E. Merck, Darmstadt, Germany).

Tabla 3.7. Composición en ácidos grasos de los colectivos de estandarización A y B así como del colectivo de validación E.

	Colectivo de Estandarización A (n=1)	Colectivo de Estandarización B (n=20)			Colectivo de Validación E (n=10)		
		Media	DT	Intervalo	Media	DT	Intervalo
C16:0 (%)	20,40	21,07	1,93	17,90-24,60	21,12	2,18	18,40-24,40
C18:0 (%)	10,00	10,41	1,80	7,70-14,90	10,51	1,43	9,00-12,60
C18:1 (%)	52,60	52,34	3,06	46,30-57,70	52,53	3,36	47,80-57,00
C18:2 (%)	10,40	9,76	1,90	6,90-13,50	9,32	1,19	8,20-12,20

Para evaluar el proceso de estandarización y transferencia de ecuaciones entre instrumentos se ha empleado un colectivo de validación E compuesto por 10 muestras no selladas de grasa de cerdo Ibérico. Estas muestras disponían de densidades ópticas (Figura 3.4) y valores de composición en ácidos grasos comprendidos dentro de los rangos de variabilidad de las muestras del colectivo de calibración y del colectivo de estandarización B (Tablas 3.6 y 3.7).

Figura 3.4. Comparación de los valores máximos y mínimos de $\log(1/R)$ para los espectros del colectivo de estandarización B y el colectivo de validación E.



Todas las muestras de grasa de cerdo Ibérico estaban conservadas originariamente en tubos de plástico de 10 ml de capacidad, envasados bajo congelación y al vacío en bolsas de plástico. Se procedió a su descongelación y fusión a 45°C para, a continuación, separar 3 alícuotas de 3 ml cada una en tubos de plástico de 5 ml de capacidad. Cada una de las 3 alícuotas iba destinada a cada uno de los laboratorios implicados en la estandarización (Córdoba, Zaragoza y Cabrils) de forma que en cada laboratorio se abriera y analizara únicamente su alícuota correspondiente.

En el caso de las muestras de grasa del colectivo A, C y D se dispuso, para cada una de ellas, de 1 único tubo de plástico de 5 ml de capacidad. Dichos tubos fueron abiertos y analizados en cada uno de los laboratorios implicados.

Para el transporte de las muestras a los diferentes laboratorios implicados en la estandarización, se procedió al envasado al vacío de las muestras correspondientes y se empleó un envase hermético con acumuladores de hielo en su interior con el fin de lograr una baja temperatura en dicho envase.

3.5.3 OBTENCIÓN DE ESPECTROS PARA MUESTRAS DE ESTANDARIZACIÓN Y TRATAMIENTO DE DATOS

Los espectros se obtuvieron, en el rango de 400 a 2.500 nm con intervalos de 2 nm, mediante los siguientes espectrofotómetros:

- Foss NIRSystems 6500 SYI equipado con un módulo de giro y ubicado en la Unidad NIR/MIR del Servicio Central de Apoyo a la Investigación de la Universidad de Córdoba.
- Foss NIRSystems 6500 SYII equipado con un módulo de transporte y ubicado en la Unidad NIR/MIR del Servicio Central de Apoyo a la Investigación de la Universidad de Córdoba.
- Foss NIRSystems 6500 SYII equipado con un módulo de giro y ubicado en Zaragoza, dentro del Departamento de Producción Animal del Servicio de Investigación Agraria de la Diputación General de Aragón.
- Foss NIRSystems 6500 SYI equipado con un módulo de transporte y ubicado en el Laboratorio Agrario de Cabrils de Consejería de Agricultura, Ganadería y Pesca de la Generalidad de Cataluña.

Antes de obtener en cada instrumento los espectros de las muestras empleadas en la estandarización, se realizó una evaluación previa del instrumento (diagnósticos o controles del instrumento) mediante la medida de la repetibilidad fotométrica o ruido, precisión de la longitud de onda y respuesta del instrumento o ganancia de los detectores.

Cada una de las muestras de los sets de estandarización (Set A a D) y validación E fueron analizadas por NIRS mediante la obtención de un espectro medio de dos submuestras, empleando para ello la misma cápsula que durante el desarrollo de las ecuaciones NIRS de predicción de ácidos grasos.

En el caso de los instrumentos localizados en el mismo laboratorio, las dos submuestras fueron analizadas en primer lugar en el instrumento Master y analizadas tras ello y de forma inmediata en el instrumento Satélite.

Para los instrumentos localizados en diferentes laboratorios, los espectros de las dos submuestras se obtuvieron en la misma semana que las muestras fueron analizadas en el instrumento Master.

Una vez se obtuvieron todos los espectros duplicados de cada una de las muestras en todos los instrumentos, se calcularon los valores de *RMS* para poder

confirmar la representatividad de ambos espectros de las muestras analizadas. De acuerdo con los resultados descritos en el apartado 4.2.1, los valores de *RMS* para espectros duplicados de grasa de cerdo Ibérico analizada mediante la modalidad de doble transmisión han de ser inferiores, aproximadamente, a $4.000 \mu\log(1/R)$. Los espectros con valores de *RMS* superiores a dicho límite fueron eliminados y la muestra analizada de nuevo hasta que sus espectros duplicados tuvieran valores de *RMS* inferiores a dicho límite.

Tanto la obtención de los espectros de las muestras empleadas en la estandarización como el tratamiento matemático de dichos datos espectrales se realizó mediante el programa ISI 3 versión 3.11 (Infrasoft International, Port Matilda, PA, USA).

Todos aquellos espectros obtenidos en el instrumento Master fueron comparados con los sus correspondientes espectros tomados en los instrumentos Satélite mediante los subprogramas CLONE y CLONE 1 incluidos en el programa ISI de acuerdo con la metodología descrita por Shenk y Westerhaus (1990, 1991a y b, 1995a y b) y Shenk *et al.* (1992). A partir de dicho tratamiento, se obtuvieron para cada instrumento 4 ficheros de estandarización (STD_{Set A}; STD_{Set B}; STD_{Set C}; STD_{Set D}) a partir de los correspondientes colectivos de estandarización (Set A, Set B, Set C and Set D). Dichos ficheros se aplicaron sobre los espectros del colectivo de validación E obtenidos en los instrumentos Satélite con el objetivo de lograr que estos espectros fueran los más parecidos posible a los espectros del colectivo de validación obtenidos en el Master.

Los resultados de estos procesos de estandarización y transferencia de ecuaciones fueron evaluados mediante dos test.

Un primer test fue la comparación del valor medio de la raíz cuadrada de las diferencias, corregidas por el sesgo, (*RMS(C)*) entre espectros de las mismas muestras obtenidos en los instrumentos Master y Satélite o entre submuestras analizadas en el mismo instrumento, a *n* longitudes de onda. La fórmula de este estadístico es:

$$RMS(C) = 10^6 \times \sqrt{\frac{\sum D^2 - \frac{\sum D^2}{n}}{n-1}} \quad D = y_a - y_b$$

donde y_a e y_b son los valores de $\text{Log}(1/R)$ de 2 espectros de la misma muestra a una longitud de onda dada. Dados los resultados descritos en el apartado 4.2.3, donde se muestra la inestabilidad del rango del Visible en cuanto al efecto de conservación de muestras bajo congelación, se decidió emplear únicamente la región NIR (1100 a 2500 nm) en el proceso de estandarización entre instrumentos y en la transferencia de las ecuaciones de calibración generadas.

Una vez los instrumentos superaron correctamente el proceso de estandarización, se aplicó un segundo test para evaluar la precisión de las ecuaciones de calibración para ácidos grasos transferidas entre instrumentos. Ello se realiza mediante la predicción del contenido en ácidos grasos de los espectros pertenecientes a las muestras del colectivo de validación E, obtenidos tanto en el instrumento Master como los Satélites previa estandarización y tras estandarizar con los diferentes ficheros de estandarización (STD_{Set A}; STD_{Set B}; STD_{Set C}; STD_{Set D}) generados para cada instrumento.

Shenk *et al.*, (1992) recomienda varios estadísticos y límites de control para evaluar las diferencias entre los valores predichos en los instrumentos Master y Satélite. En opinión de Shenk *et al.*, (1992), uno de los estadísticos más importantes es el error típico de las diferencias (*ETD*) o error típico de las diferencias entre los valores predichos para N muestras analizadas en el instrumento Master y Satélite. El valor de *ETD* se calcula de acuerdo con la siguiente fórmula:

$$ETD = \sqrt{\frac{\sum (\hat{Y}_{Master} - \hat{Y}_{Satélite})^2}{N}}$$

donde \hat{Y}_{Master} e $\hat{Y}_{Satélite}$ son los valores predichos para una muestra en los instrumentos Master y Satélite respectivamente.

Independientemente de este estadístico, otro estadístico de gran utilidad para evaluar los resultados del proceso de estandarización y transferencia de ecuaciones es el valor medio del estadístico H de Mahalanobis del colectivo de validación. Este estadístico aporta información acerca de cómo de distantes se encuentran las muestras predichas por el Satélite de aquellas predichas por el Master. Si el valor obtenido de H es superior a 3 se considera que los espectros de una muestra o población se encuentran alejado de aquellos de las muestras del colectivo de calibración obtenidos en el Master

y, por tanto, ambos colectivos espectrales son muy diferentes (Shenk *et al.*, 1992; Shenk y Westerhaus, 1995a y b; Park *et al.*, 1999).

4 RESULTADOS Y DISCUSIÓN

El presente capítulo se ha estructurado en una serie de apartados que corresponden a estudios específicos diseñados para tratar de responder a los objetivos marcados en esta Tesis Doctoral. Para entender en toda su magnitud la estructuración empleada, es preciso también conocer algunos aspectos ligados al plan de trabajo y diseño experimental seguido. Así, si bien el objetivo general y los específicos eran claros desde la realización de la presente Tesis Doctoral, el plan de trabajo y diseño experimental ha sido necesario adaptarlo en función de los resultados preliminares obtenidos.

Era un compromiso del doctorando, de sus Directores y del Grupo de Investigación del Dpto. de Producción Animal de la E.T.S.I.A.M. de la Universidad de Córdoba mostrar al sector productor de cerdo Ibérico la viabilidad de la tecnología NIRS para el análisis de ácidos grasos en grasa de cerdo. Para ello era necesario no sólo el profundizar en aspectos científicos básicos de la tecnología NIRS sino también en aspectos metodológicos relacionados con la normativa oficial de clasificación de canales de cerdo Ibérico en función del régimen alimenticio durante el cebo.

Por ello, en el presente apartado 4.1 se engloban los estudios relacionados con la metodología oficial de toma de muestras y análisis de grasa de cerdo Ibérico, en el apartado 4.2 se realiza la puesta a punto de la tecnología NIRS para la recogida de datos espectroscópicos de grasa fundida y tejido adiposo subcutáneo y en el apartado 4.3 a modo de análisis de viabilidad se desarrollan ecuaciones preliminares para la predicción de ácidos grasos en grasa fundida para profundizar con posterioridad en la minimización de las fuentes de error que afectan a la capacidad predictiva de dichas ecuaciones.

En el apartado 4.4 se desarrollan modelos de clasificación o autenticación de canales de cerdo Ibérico basados en el uso como variables de la composición en ácidos grasos, información obtenida mediante cromatografía de gases y datos predichos NIRS, junto con la información espectral NIR *per se*. Posteriormente, en el apartado 4.5, se realiza la estandarización de instrumentos NIRS junto con la transferencia de ecuaciones de grasa fundida y la evaluación de la repetibilidad intra e interlaboratorios de la determinación, mediante la técnica NIRS, del contenido en ácidos grasos en grasa de cerdo Ibérico.

4.1 OPTIMIZACIÓN DE LA METODOLOGÍA DE ANÁLISIS DE ÁCIDOS GRASOS EN CANALES DE CERDO IBÉRICO

4.1.1 ESTUDIO DE LOCALIZACIÓN ANATÓMICA DE TOMA DE MUESTRAS

El objetivo de este estudio es el de evaluar las diferencias en cuanto a composición de ácidos grasos en muestras de grasa subcutánea obtenidas en diferentes localizaciones anatómicas de la canal.

En la Tabla 4.1.1 se recoge la composición media en ácidos grasos de la grasa subcutánea de la canal en las tres zonas estudiadas (Figura 3.1). Como puede apreciarse, la composición total en ácidos grasos saturados fue superior en la región B (grasa de jamón, entre la inserción del rabo y la corva) mientras que la suma de ácidos grasos insaturados fue más alta en la región A (grasa subcutánea de la zona de inserción del rabo en el jamón). La composición media de ácidos grasos mayoritarios de la región C (grasa de la zona de inserción del rabo con la zona coxal) presentó valores intermedios entre la composición de las muestras de las zonas A y B.

En la Tabla 4.1.1 se incluyen los resultados obtenidos tras realizar el análisis de la varianza sobre la composición media de las diferentes zonas. Como se observa en dicha tabla, se detectaron diferencias significativas, al nivel del 1%, en la composición de alguno de los ácidos grasos existentes en las tres zonas anatómicas. Dichas diferencias fueron evidentes para los ácidos grasos C18:0 y C20:1, siendo más difusas

en la mayoría de ácidos grasos (C16:0, C16:1, C17:1, C18:1, C18:2 y C20:0) e incluso no se observaron diferencias significativas entre los ácidos grasos minoritarios C14:0, C17:0 y C18:3. El ácido graso C18:2 constituye una excepción dentro de los ácidos grasos mayoritarios, ya que no se detectaron diferencias significativas entre la composición de las zonas A y C pero sí entre ambas y la existente en la zona B.

Tabla 4.1.1. Composición media en ácidos grasos de grasa subcutánea de cerdos ibéricos, según la zona de toma de muestras para el análisis*

Ácido graso	Zona Anatómica					
	A		B		C	
	Media	DT	Media	DT	Media	DT
C14:0 (%)	1,41	0,18	1,41	0,19	1,32	0,15
C16:0 (%)	22,36 a	1,19	24,41 b	1,24	21,88 a	0,80
C16:1 (%)	2,57 a	0,42	2,72 a	0,43	2,03 b	0,22
C17:0 (%)	0,33	0,07	0,33	0,06	0,36	0,06
C17:1 (%)	0,35 a	0,06	0,29 b	0,06	0,34 a	0,06
C18:0 (%)	9,72 a	1,02	11,80 b	1,28	10,73 c	0,85
C18:1 (%)	53,24 a	1,70	49,80 b	1,45	52,76 a	1,13
C18:2 (%)	7,42 ab	0,51	7,07 a	0,82	7,83 b	0,45
C18:3 (%)	0,46	0,07	0,51	0,13	0,45	0,07
C20:0 (%)	0,20 ab	0,04	0,18 a	0,04	0,21 b	0,03
C20:1 (%)	1,62 a	0,26	1,24 b	0,24	1,81 c	0,26

*Valores medios con letras diferentes en la misma línea muestra diferencias significativas al nivel del 1%.

Respecto a los valores de desviaciones típicas de los diferentes ácidos grasos, se observa en dicha tabla que los valores mínimos se obtuvieron para todos los ácidos grasos (excepto para el C20:1) en la zona anatómica C. Ello muestra la menor variabilidad en cuanto a composición de la grasa de dicha zona anatómica frente a las otras zonas analizadas.

Teniendo en cuenta los resultados obtenidos se recomienda que la toma de muestras de grasa para el análisis de ácidos grasos se realice preferentemente en la zona C. A la elección de dicha zona anatómica también contribuye el hecho de la mayor facilidad de toma de muestra de la grasa en la línea de sacrificio en dicha zona, frente a las zonas A y B.

4.1.2 ESTUDIO DEL NÚMERO MÍNIMO DE MUESTRAS REPRESENTATIVAS

Con el objetivo de estudiar el número mínimo de muestras procedentes de animales individuales de un determinado lote, necesarias para obtener una muestra representativa de dicho lote, se ensayaron diferentes porcentajes de animales (10, 20, 30 o 40%) procedentes de lotes con diferente alimentación durante la etapa de cebo (apartado 3.1.2).

En la Tabla 4.1.2 se recoge, para cada lote experimental, el número de animales que componía dicho lote junto con los valores medios, máximos y mínimos de los 4 ácidos grasos mayoritarios (C16:0, C18:0, C18:1, C18:2). Como puede apreciarse en dicha tabla, con los lotes experimentales considerados se cubre gran parte de la variabilidad existente en cuanto a composición de ácidos grasos en grasa de cerdo Ibérico se refiere (De Pedro, 1995 y 2000).

Tabla 4.1.2. Número de animales, valores medios, mínimos y máximos de composición en ácidos grasos mayoritarios en grasa subcutánea de cerdo Ibérico de cada lote experimental y su clasificación comercial.

LOTE	N	C16:0 (%)			C18:0 (%)			C18:1 (%)			C18:2 (%)			C*
		Med	Min	Max	Med	Min	Max	Med	Min	Max	Med	Min	Max	
I	37	19,7	17,8	21,8	9,2	6,9	11,1	56,4	53,4	59,4	8,7	7,9	9,9	B
II	32	20,3	18,5	23,4	9,1	7,4	10,6	54,6	50,6	56,6	9,4	8,2	10,9	B
III	80	20,4	18,5	22,7	9,7	8,3	11,8	53,8	51,2	56,4	9,7	8,3	11,5	R
IV	36	22,4	19,3	24,1	10,3	7,1	13,5	52,5	50,0	58,7	8,0	6,6	9,9	R
V	45	20,9	19,6	22,5	10,4	8,9	12,0	53,6	51,9	55,2	8,8	7,8	9,6	R
VI	71	21,3	19,2	23,0	11,5	9,7	13,2	51,9	49,6	54,4	8,9	8,0	10,2	P

* Clasificación de la composición media de acuerdo con los criterios del contrato tipo homologado (B: Bellota; R: Recebo; P:Piense).

La Tabla 4.1.2 incluye también la clasificación de cada lote de acuerdo con los criterios de calidad para composición en ácidos grasos existente en el contrato tipo homologado para la campaña 98/99 (BOE, 1998), abarcando las 3 categorías comerciales existentes (*bellota, recebo y pienso*).

En las Tablas 4.1.3 a 4.1.8 se muestran los resultados obtenidos al calcular, en las 5 diferentes series de selección de animales ensayadas, la composición media de las muestras elegidas al azar en cada lote de animales, según el porcentaje de muestras empleado para el cálculo.

En dichas tablas se observa, en primer lugar, cómo dependiendo del número de muestras tomadas al azar en un lote, su clasificación analítica de acuerdo con los criterios de contrato tipo homologado podría no coincidir con el del lote en su conjunto (Tabla 4.1.2). Así, en el Lote I, de tipo *bellota*, según los valores medios de los 37 animales que la componían (Tabla 4.1.2), bastaría una toma de muestras al azar del 20 % para que la composición en ácidos grasos de esas muestras permitiera clasificar la partida en la misma categoría que si se tomase la totalidad de las muestras (Tabla 4.1.3). Si tomásemos sólo un 10 % de muestras correríamos el riesgo de que en alguna ocasión (como se observa en la serie 5 de selección de muestras al azar) la partida fuese clasificada de la categoría *recebo* en lugar de *bellota* (Tabla 4.1.3).

En el caso del lote II, también de la categoría *bellota* según la media de las 32 muestras de los animales que componían el lote (Tabla 4.1.2), es preciso llegar a un 40 % de muestras tomadas al azar para tener la garantía de que la composición en ácidos grasos de esas muestras van a permitir clasificar la partida en la misma categoría que se analizaran las 32 muestras del lote. Tomar un 20 o un 30 % de muestras supondría dar una clasificación errónea de la partida en algunas ocasiones (Tabla 4.1.4).

En el caso del lote III (Tabla 4.1.5), se aprecia que en todos los supuestos de selección de muestras, el resultado de su composición media daría lugar a una clasificación similar (*recebo*) a la obtenida con el total de las muestras de la partida (Tabla 4.1.2). Sin embargo, no ocurre lo mismo con los lotes IV y V (Tabla 4.1.6 y 4.1.7) en los cuales, aun tomando un 40% de muestras, en algunos muestreos el resultado de la composición media sería tal que la partida se clasificaría en la categoría de *pienso*, cuando la media de toda la partida posee un perfil de ácidos grasos que se encuadra en la categoría de *recebo* (Tabla 4.1.2).

Tabla 4.1.3. Valores medios del contenido en ácidos grasos (%) de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote I (n=37).

%	Muestras	SERIE																								
		1					2					3					4					5				
		16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C
10	4	18,7	8,9	57,8	8,6	B	19,7	9,5	56,3	8,3	B	20,2	9,4	56,5	8,1	B	19,5	9,2	56,7	8,7	B	20,0	9,6	55,7	8,9	R
20	7	18,9	8,7	57,6	8,7	B	19,8	9,4	56,1	8,6	B	20,0	9,4	56,5	8,2	B	19,9	9,4	56,0	8,6	B	19,7	9,5	55,9	8,8	B
30	11	19,3	9,0	57,0	8,7	B	19,8	9,5	55,9	8,6	B	19,8	9,4	56,6	8,3	B	19,7	9,1	56,4	8,6	B	19,4	9,3	56,3	8,9	B
40	14	19,5	9,0	56,8	8,7	B	19,8	9,5	55,9	8,7	B	19,7	9,2	56,6	8,4	B	19,5	9,2	56,5	8,6	B	19,6	9,4	56,1	8,8	B

Tabla 4.1.4. Valores medios del contenido en ácidos grasos (%) de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote II (n=32).

%	Muestras	SERIE																								
		1					2					3					4					5				
		16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C
10	3	19,3	8,1	55,5	10,0	R	21,3	10,0	53,1	9,3	R	20,5	9,3	54,7	9,2	B	19,4	8,7	55,9	9,4	B	19,5	9,6	55,6	9,6	R
20	6	20,1	8,3	54,7	9,8	R	20,4	9,3	54,3	9,5	B	20,3	9,1	55,0	9,2	B	20,3	9,0	54,6	9,4	B	19,7	9,7	55,0	9,7	R
30	10	20,3	8,8	54,5	9,6	R	20,2	9,2	54,6	9,5	B	20,6	9,3	54,5	9,2	B	20,3	9,2	54,4	9,4	B	20,3	9,4	54,4	9,4	B
40	13	20,3	9,0	54,6	9,5	B	20,6	9,2	54,2	9,4	B	20,4	9,2	54,6	9,3	B	20,3	9,3	54,4	9,4	B	20,3	9,4	54,4	9,4	B

C: Clasificación; B: Bellota; R: Recebo; P:Pienso

Tabla 4.1.5. Valores medios del contenido en ácidos grasos (%) de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote III (n=80).

%	Muestras	SERIE																								
		1					2					3					4					5				
		16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C
10	8	20,5	9,8	53,6	9,8	R	20,3	9,9	53,8	9,6	R	20,3	9,7	54,2	9,6	R	20,3	9,6	53,8	9,9	R	20,7	10,2	53,2	9,6	R
20	16	20,5	9,7	53,7	9,8	R	20,4	9,8	53,7	9,7	R	20,4	9,8	53,7	9,7	R	20,5	9,8	53,5	9,8	R	20,6	9,8	53,5	9,7	R
30	24	20,5	9,6	53,7	9,8	R	20,4	9,7	53,8	9,7	R	20,4	9,7	53,8	9,7	R	20,4	9,8	53,7	9,7	R	20,5	9,7	53,7	9,7	R
40	32	20,5	9,7	53,6	9,8	R	20,4	9,8	53,7	9,7	R	20,5	9,7	53,8	9,6	R	20,4	9,7	53,8	9,7	R	20,5	9,7	53,6	9,7	R

Tabla 4.1.6. Valores medios del contenido en ácidos grasos (%) de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote IV (n=36).

%	Muestras	SERIE																								
		1					2					3					4					5				
		16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C
10	4	22,7	10,7	52,1	7,9	P	23,3	11,5	51,3	7,2	P	22,8	10,7	51,9	7,7	P	22,7	10,9	51,5	8,0	P	21,6	9,4	53,3	8,9	R
20	7	22,3	10,1	52,9	7,9	R	22,6	10,8	51,9	7,9	P	22,9	10,7	51,7	7,8	P	22,6	10,8	51,8	7,9	P	22,3	10,2	52,3	8,2	R
30	11	22,1	10,0	53,3	7,9	R	22,9	11,2	51,6	7,6	P	22,8	10,8	51,7	7,8	P	22,7	10,7	51,9	7,9	P	22,1	10,0	52,9	8,3	R
40	14	22,3	10,2	53	7,8	R	22,7	10,9	51,8	7,8	P	22,9	10,8	51,7	7,8	P	22,5	10,6	52,2	8,0	P	22,3	10,4	52,5	8,0	R

C: Clasificación; B: Bellota; R: Recebo; P:Pienso

Tabla 4.1.7. Valores medios del contenido en ácidos grasos (%) de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote V (n=45).

%	Muestras	SERIE																								
		1					2					3					4					5				
		16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C
10	5	20,9	10,1	53,9	8,6	R	20,8	9,7	53,8	9,1	R	20,6	10,3	53,8	8,9	R	20,8	10,5	53,5	9,0	R	20,7	9,7	53,9	9,2	R
20	9	21,1	10,4	53,6	8,5	R	20,9	10,0	53,6	9,0	R	20,7	10,4	53,7	8,9	R	20,7	10,2	53,7	9,1	R	20,9	10,4	53,5	9,0	R
30	14	20,9	10,5	53,7	8,7	R	20,8	10,0	53,7	9,0	R	20,6	10,6	53,6	9,0	P	20,6	10,3	53,7	9,1	R	21,0	10,5	53,4	8,8	R
40	18	20,9	10,4	53,7	8,7	R	20,9	10,2	53,6	8,9	R	20,7	10,6	53,6	8,9	P	20,6	10,2	53,8	9,0	R	20,9	10,6	53,4	8,9	P

Tabla 4.1.8. Valores medios del contenido en ácidos grasos (%) de las muestras de grasa, elegidas al azar en cinco series, según el porcentaje de muestras tomadas en el lote VI (n=71).

%	Muestras	SERIE																								
		1					2					3					4					5				
		16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C	16:0	18:0	18:1	18:2	C
10	7	21,4	11,6	51,6	9,0	P	21,2	11,9	51,6	9,0	P	21,0	11,9	51,7	9,1	P	21,4	11,2	51,9	9,0	P	21,2	11,8	51,8	9,0	P
20	14	21,4	11,5	51,5	9,1	P	21,5	11,9	51,4	8,8	P	21,3	11,4	51,9	9,1	P	21,4	11,3	51,9	8,9	P	21,5	11,6	51,6	8,9	P
30	21	21,2	11,4	51,8	9,1	P	21,5	11,7	51,6	8,8	P	21,5	11,7	51,6	8,8	P	21,3	11,3	51,8	9,1	P	21,4	11,6	51,8	8,9	P
40	28	21,3	11,5	51,7	9,0	P	21,4	11,6	51,7	8,9	P	21,2	11,5	51,9	9,0	P	21,3	11,4	51,8	9,0	P	21,4	11,6	51,6	8,9	P

C: Clasificación; B: Bellota; R: Recebo; P: Pienso

Un caso menos problemático sería el del lote VI (Tabla 4.1.8) en el que, cualquiera que fuese el porcentaje de muestras tomado de la partida, el resultado de su clasificación analítica sería coincidente con el de la media del total de muestras de la partida (Tabla 4.1.2). Puede que este hecho se deba probablemente a que los niveles de ácido esteárico son muy elevados en la mayoría de las muestras del lote, lo que fácilmente conlleva a su clasificación en la categoría *pienso*.

Por tanto, se observa la importancia que tiene el modo en el que se realiza el muestreo, ya que en algunos casos sería suficiente un 10 o 20 % de muestras de la partida tomadas al azar, mientras que en otros, incluso llegando a tomar el 40 % de las muestras para obtener una muestra representativa, su composición media no refleja la calidad que corresponde al total del lote. Incluso la clasificación parece estar afectada por los individuos seleccionados para un porcentaje dado.

Cabe preguntarse si, incluso tomando el 100% de muestras (es decir, si se tomaran muestras de todos los animales que componen cada lote) y analizando una única muestra media, se estaría perdiendo información valiosa para el ganadero, para el industrial y para el consumidor. Así, parece probable que exista una importante variabilidad a nivel de individuos de tal forma que, aunque la media de un lote indicara su clasificación en la categoría *recebo*, en dicho lote podrían existir animales con perfiles de ácidos grasos correspondientes a la categoría *bellota*, *recebo* o incluso *pienso* (De Pedro, 2000).

Dado que se disponía de la composición en ácidos grasos de cada uno de los animales pertenecientes a cada lote, se aplicaron los criterios de calidad definidos por el contrato tipo homologado a las muestras de grasa de cada uno de los animales que componían los lotes estudiados y se compararon estos resultados con la clasificación a partir de la muestra media de cada lote.

En la Tabla 4.1.9 se recogen, para los lotes I a VI, los resultados de clasificación de las canales a partir de la muestra media y la clasificación obtenida a partir de la composición de las muestras de grasa de los animales de cada lote.

En dicha tabla se observa que, de acuerdo con la composición media en ácidos grasos de los lotes, (Tabla 4.1.2) y según los criterios analíticos del contrato tipo

homologado, los lotes I y II se considerarían de la categoría comercial *bellota*, los tres siguientes lotes (III, IV y V) de tipo *recebo* y el último de tipo *pienso*.

Tabla 4.1.9. Clasificación de cada lote experimental, en función de su composición media y número de animales clasificados en cada categoría, al aplicar los criterios del CTH de forma individualizada.

Lote	Clasificación del lote	Número de muestras de cada categoría		
		B	R	P
I (n=37)	B	24	11	2
II (n=32)	B	15	12	5
III (n=80)	R	11	51	18
IV (n=36)	R	4	12	20
V (n=45)	R	5	22	18
VI (n=71)	P		7	64

B: Bellota; R: Recebo; P: Pienso

Al aplicar los criterios de calidad analítica del contrato tipo a los individuos de cada lote se obtiene que en los lotes I y II, de categoría *bellota*, la mayor parte de las canales serían, según su composición en ácidos grasos, de tipo *bellota* (24/37 y 15/32 respectivamente), otras serían “recebo” (11/37 y 12/32 respectivamente) e incluso “pienso” (2/37 y 5/32 respectivamente). En los lotes III, IV y V ocurre algo similar y tan sólo en el caso del lote VI, sólo 7 muestras serían considerado de tipo *recebo* y 64 (de un total de 71) serían consideradas de tipo *pienso*, al igual que la media de dicho lote.

Por tanto se aprecia que, incluso considerando la muestra media de la totalidad de animales que componen un lote, la clasificación que se realiza de esa partida en función de la composición en ácidos grasos de una muestra media, no reflejaría la variabilidad existente en cuanto a composición, y por tanto en la clasificación comercial, de las canales individuales que componen un lote.

Los resultados obtenidos en el presente estudio son de enorme importancia en cuanto a la obtención de conclusiones respecto a la utilización de la composición en ácidos grasos de la grasa subcutánea como criterio para certificar el régimen alimenticio de los cerdos Ibéricos en el período de cebo de acuerdo con los criterios reflejados en el contrato tipo homologado.

De utilizarse dicho sistema de evaluación del régimen alimenticio, para una correcta clasificación de lotes de cerdos Ibéricos sería imprescindible el realizar un análisis individual de cada uno de los animales que compongan dicho lote, máxima teniendo en cuenta el diferente precio de la canal dependiendo de la categoría asignada. Esta errónea clasificación puede llevar consigo un incremento importante en el coste para el industrial (o una disminución del beneficio para el ganadero) lo que finalmente encarecería el precio final de los productos acabados.

Asimismo, la necesidad de analizar la composición en ácidos grasos de cada canal perteneciente a un mismo lote implicaría mayor retraso en el pago a los ganaderos, al incrementarse sensiblemente el tiempo de análisis necesario para la obtención de la información cromatográfica de la totalidad de animales sacrificados.

Los resultados de este estudio refuerzan la necesidad de puesta a punto de métodos de análisis rápidos y económicos que permitieran ampliar el control de calidad que actualmente se realiza a nivel de los organismos de control e industrias del sector del cerdo Ibérico.

4.1.3 ENSAYO COLABORATIVO

Este ensayo colaborativo, descrito en el apartado 3.1.3 y en el que participaron 6 laboratorios relevantes del sector, se diseñó para determinar la repetibilidad y reproducibilidad existente en el análisis de ácidos grasos en grasa de cerdo Ibérico mediante cromatografía de gases.

El conocimiento del error intra e interlaboratorios es de gran importancia por diferentes aspectos pero, fundamentalmente y dentro del contexto de esta Tesis Doctoral, para evaluar con posterioridad la precisión y exactitud de las calibraciones NIRS generadas.

Tras el análisis de las muestras de grasa de cerdo Ibérico de acuerdo con el protocolo descrito en el apartado 3.1.3, los laboratorios implicados en este ensayo colaborativo aportaron los valores del perfil completo de ácidos grasos existentes en la grasa (desde C14:0 a C20:1). Sin embargo, sólo se consideraron en este estudio los

valores correspondientes a los ácidos grasos mayoritarios (con concentración superior al 5%) en dicha grasa como son C16:0, C18:0, C18:1 y C18:2.

El motivo fundamental de dicha simplificación fue el considerar para los cálculos de este estudio, únicamente aquellos 4 ácidos grasos empleados por el contrato tipo homologado y por la mayoría de las Denominaciones de Origen para clasificar los cerdos en función del tipo de alimentación. El resto de ácidos grasos minoritarios (todos ellos suman una concentración aproximada del 6%) no son tenidos en cuenta por el Sector para la clasificación de las muestras y tampoco se consideraron en este estudio.

De acuerdo con las recomendaciones realizadas, cada uno de los 6 laboratorios participantes en el estudio debía de realizar el análisis de cada una de las muestras remitidas por duplicado, con el objetivo de poder evaluar el error existente dentro de cada laboratorio (repetibilidad o error intralaboratorio). Sin embargo, 2 laboratorios (Lab 2 y Lab 4) sólo aportaron 1 único dato por muestra. Si bien esto complicó el tratamiento estadístico de los datos, ningún laboratorio fue rechazado a priori y todos los resultados se incluyeron en los cálculos del estudio. Para resolver este inconveniente, en el cálculo de la repetibilidad sólo se tuvieron en cuenta los datos de 4 laboratorios mientras que para el cálculo de la reproducibilidad (o errores interlaboratorios), se incluyeron los datos de los 6 laboratorios participantes.

La Tabla 4.1.10 recoge los resultados de composición media y desviación típica de los 4 ácidos grasos mayoritarios aportados por los 6 laboratorios participantes en el estudio colaborativo.

Tabla 4.1.10. Media y desviación típica de ácidos grasos en las muestras analizadas por cada laboratorio*

Ácido graso	LAB 1		LAB 2		LAB 3		LAB 4		LAB 5		LAB 6	
	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT	Media	DT
C16:0 (%)	23,02 a	1,37	23,13 a	1,35	24,69 a	1,42	24,29 b	1,33	23,39 a	1,33	23,34 a	1,38
C18:0 (%)	11,00 a	0,92	10,85 a	0,94	10,70 a	0,96	9,78 b	1,20	10,98 a	0,92	10,67 a	0,94
C18:1 (%)	49,25 a	1,74	50,45 c	1,83	48,76 a	1,88	50,15 bc	1,60	49,39 ab	1,78	50,26 bc	1,85
C18:2 (%)	9,44 ab	0,56	9,33 bc	0,55	9,11 c	0,55	9,67 a	0,69	9,45 ab	0,55	9,23 bc	0,54

*Valores medios seguidos de letras diferentes son significativamente diferentes a nivel del 1 % ($p < 0,01$)

Tras una simple comparación de los valores medios de los ácidos grasos, se observa que existen diferencias entre los datos aportados por los diferentes laboratorios. Sin embargo, dadas los altos valores de las desviaciones típicas (*DT*) de los ácidos grasos, provocados por la variabilidad animal existente dentro de una partida, para una correcta comparación de los valores medios de los ácidos grasos se procedió a realizar un análisis de la varianza junto con un test de comparación de medias.

La Tabla 4.1.10 muestra la existencia de diferencias significativas tras el análisis de la varianza de los datos medios aportados por cada laboratorio, si bien los resultados fueron diferentes dependiendo del ácido graso considerado. Así, las variables C16:0 y C18:0 destacan por una mayor homogeneidad en los valores medios, observándose diferencias en un pequeño número de laboratorios (Lab 3 y Lab 4 para el C16:0 y Lab 4 para el C18:0). Sin embargo, para los valores medios de C18:1 y C18:2, se observaron diferencias significativas entre un mayor número de laboratorios.

Cabe pensar que las diferencias entre laboratorios podrían venir motivadas por la existencia de dos diferentes métodos de extracción empleados en dichos laboratorios de la grasa fundida a partir del tejido adiposo. Por ello se compararon los valores obtenidos por los laboratorios según el método empleado para la obtención de la grasa líquida a partir del tejido adiposo ya sea fusión mediante microondas o extracción mediante disolvente orgánico.

La Tabla 4.1.11 muestra los valores medios y desviaciones típicas de los ácidos grasos analizados en los 6 laboratorios según el método de extracción de grasa líquida del tejido adiposo que se haya empleado. Como se observa, no existen diferencias significativas a nivel de valores medios entre los resultados obtenidos por ambos métodos de extracción. Estos resultados son similares a los obtenidos por De Pedro *et al.* (1996) si bien, en dicho trabajo, se determinó la composición de la grasa, por cromatografía de gases, en un único laboratorio de análisis.

Los resultados obtenidos en el presente trabajo, confirman los resultados de De Pedro *et al.* (1996) referentes a la similitud de la composición en ácidos grasos independientemente del método de obtención de grasa fundida, con la ventaja que refuerza la posibilidad de aplicación de la fusión en horno microondas al ser la fusión

más rápida, sencilla, económica y no necesitar reactivos químicos ni generar residuos contaminantes.

Tabla 4.1.11. Media y desviación típica de ácidos grasos en las muestras según el método de extracción de grasa empleado*

Ácido Graso	Extracción con disolvente		Fusión en horno microondas	
	Media	DT	Media	DT
C16:0 (%)	23,69 a	1,52	23,60 a	1,46
C18:0 (%)	10,58 a	1,13	10,75 a	0,98
C18:1 (%)	49,68 a	1,94	49,74 a	1,81
C18:2 (%)	9,44 a	0,62	9,30 a	0,57

* Nivel de significación del 1%

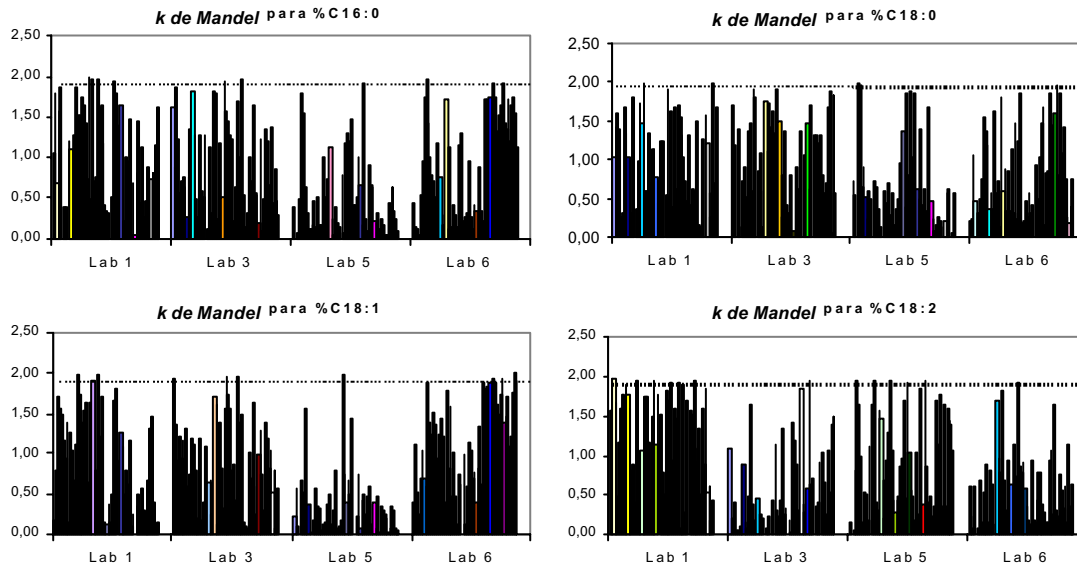
Independientemente de este resultado, los datos incluidos en la Tabla 4.1.11 confirman que las diferencias significativas entre laboratorios mostradas en la Tabla 4.1.10 son debidas exclusivamente a diferencias en la propia determinación del contenido en ácidos grasos mediante cromatografía de gases, método oficial para la clasificación de canales de cerdo Ibérico de acuerdo con los criterios del contrato tipo homologado.

La Figura 4.1.1 contiene las representaciones gráficas del estadístico k de Mandel de las muestras analizadas, agrupados por laboratorios, para cada uno de los ácidos grasos estudiados. Se ha de matizar que dicho estadístico informa acerca de la incertidumbre de los datos (error) dentro de cada laboratorio. Los valores límite de dicho estadístico para considerar una muestra como anómala (al nivel del 1%) en el caso de 4 laboratorios (son sólo 4 los laboratorios que aportan análisis por duplicado) y 2 repeticiones por muestra es 1.91 (ISO, 1994). La situación más deseable sería aquella en la cual los valores de k de Mandel para la variable y las muestras analizadas en cada laboratorio tuvieran un valor lo más próximo a 0. Se ha de tener en cuenta que dicho estadístico únicamente puede presentar valores positivos.

Del análisis de la Figura 4.1.1 se observa que los valores de k de Mandel presentan un valor elevado en la mayoría de los laboratorios y ácidos grasos estudiados. Esto indicaría que todos los laboratorios van a presentar una baja repetibilidad para todos los ácidos grasos estudiados. Sin embargo, se aprecia un número muy bajo de

muestras anómalas en cada uno de los ácidos grasos (de 1 a 7), no detectándose, de forma sistemática, por algún laboratorio en concreto.

Figura 4.1.1. Estadísticos k de Mandel, agrupados por laboratorios.



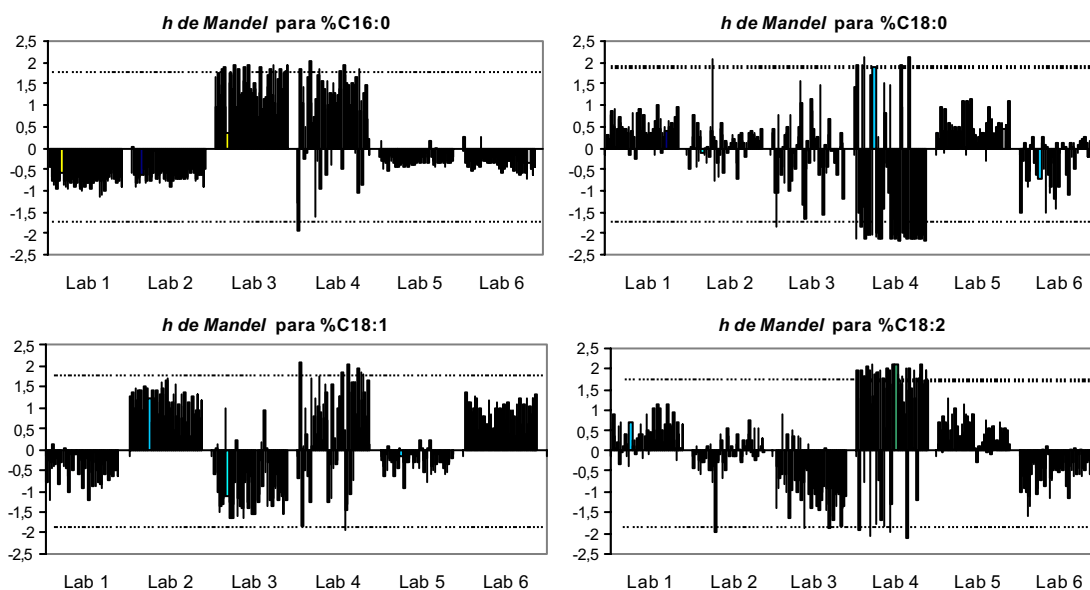
La Figura 4.1.2 contiene las representaciones gráficas del estadístico h de Mandel de las muestras analizadas, agrupados por laboratorios, para cada uno de los ácidos grasos considerados. Recordar que dicho estadístico informa acerca de la incertidumbre entre laboratorios, evaluada ésta desde el punto de vista de las diferencias entre los valores medios de las muestras. Los valores límite de dicho estadístico para considerar una muestra como anómala (al nivel del 1%) en el caso de 6 laboratorios y 2 repeticiones por muestra es $\pm 1,87$ (ISO, 1994). La situación óptima es aquella en la cual los valores de h de Mandel para la variable y muestras analizadas en cada laboratorio tuvieran un valor lo más próximo a 0, con una alternancia aleatoria en cuanto al signo que presenta.

Del análisis de la Figura 4.1.2 se observa que, para cada una de los ácidos grasos estudiados, los valores de h de Mandel siguen un comportamiento diferente en cada laboratorio. Así, para el ácido graso C16:0, los valores de h para Lab 3 y Lab 4 son excesivamente altos, aproximándose y sobrepasando en el caso de varias muestras los valores fijados como límites para la detección de datos anómalos.

Al mismo tiempo, los valores de h en Lab 3 y en la mayoría de los datos de Lab 4 presentan valores positivos mientras que el resto de laboratorios (Lab 1, Lab 2,

Lab 5 y Lab 6) presentan valores negativos. Eso indica que, para este ácido graso (C16:0), existen diferencias sistemáticas o sesgos entre los valores medios obtenidos en el primer grupo de laboratorios (Lab 3 y Lab 4) y el segundo grupo (Lab 1, Lab 2, Lab 5 y Lab 6). También se puede observar una presencia de sesgos importantes entre los 2 grupos de laboratorios para el ácido graso C18:1.

Figura 4.1.2. Estadísticos h de Mandel, agrupados por laboratorios.



En lo referente a la detección de muestras anómalas, Lab 4 presenta en general muestras anómalas en todos los ácidos grasos considerados, siendo mayor en número para las variables C18:0 y C18:2. Esto nos indicaría que, en un número importante de muestras, Lab 4 presenta diferencias excesivamente altas (significativas) frente al resto de laboratorios. Otros laboratorios que presentan un cierto número de muestras anómalas son Lab 3 (fundamentalmente en la variable C16:0) y Lab 2.

Por tanto, tras la representación gráfica de los estadísticos h y k de Mandel se obtuvo una primera evaluación de las diferencias existentes (tanto intra como entre laboratorios) y de la procedencia de las muestras detectadas como anómalas (de forma importante en Lab 4).

Sin embargo, y de acuerdo con la norma ISO 5725-2 (1994), estas representaciones gráficas son una herramienta para obtener una primera valoración de los resultados pero no han de ser empleadas en una toma de decisiones drástica. Es por

ello que los datos de las muestras que fueron detectadas como anómalas por ambos test gráficos no fueron eliminados tras este análisis.

A continuación se procedió al cálculo de los test numéricos empleados para la detección de muestras anómalas (test de Cochran y test de Grubbs simple y doble).

La Tabla 4.1.12 muestra el número de muestras detectadas como anómalas en cada laboratorio y para cada una de las variables tras realizar el test de Cochran.

En ella se observa que Lab 1 es el laboratorio que presenta un mayor número de muestras anómalas para todos los ácidos grasos considerados, es decir, la variabilidad que presenta los duplicados en algunas muestras determinadas en este laboratorio es superior a lo admisible. Esto podría explicarse porque Lab 1 fue el único laboratorio que realizó una inyección manual de las muestras en el cromatógrafo (el resto de laboratorios empleó inyector automático) lo cual suele considerarse como una fuente de error importante en el análisis de réplicas (Salas, comunicación personal).

Tabla 4.1.12. Muestras anómalas detectadas tras el Test de Cochran

Acido graso	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6
C16:0	2	n.d.	0	n.d.	0	0
C18:0	2	n.d.	0	n.d.	1	0
C18:1	1	n.d.	0	n.d.	0	1
C18:2	2	n.d.	0	n.d.	0	0

n.d.: dato no disponible

La Tabla 4.1.13 muestra el número de muestras detectadas como anómalas en cada laboratorio y por variable, tras realizar el test de Grubbs sencillo y doble.

Tabla 4.1.13. Muestras anómalas detectadas tras el Test de Grubbs simple y doble

Ácido graso	Lab 1		Lab 2		Lab 3		Lab 4		Lab 5		Lab 6	
	Sencillo	Doble	Sencillo	Doble	Sencillo	Doble	Sencillo	Doble	Sencillo	Doble	Sencillo	Doble
C16:0	0	0	0	0	5	2	0	2	0	0	0	0
C18:0	0	0	0	0	0	2	19	2	0	0	0	0
C18:1	0	0	0	0	0	0	0	0	0	0	0	0
C18:2	0	0	0	0	1	1	4	0	0	0	0	1

Como se observa, el mayor número de anómalos se observa para Lab 4 (fundamentalmente para C18:0 y en menor medida para C16:0 y C18:2) y Lab 3 (para C16:0 y en menor medida para C18:0 y C18:2). Esto indica que, para dichas muestras, los valores de los ácidos grasos determinados por Lab 3 y Lab 4, presentan valores que difieren estadísticamente de los datos aportados por el resto de laboratorios participantes.

Todos los valores indicados de las muestras detectadas como anómalas tanto por los test de Cochran como de Grubbs (sencillo y doble) no se consideraron en los cálculos posteriores de repetibilidad y reproducibilidad.

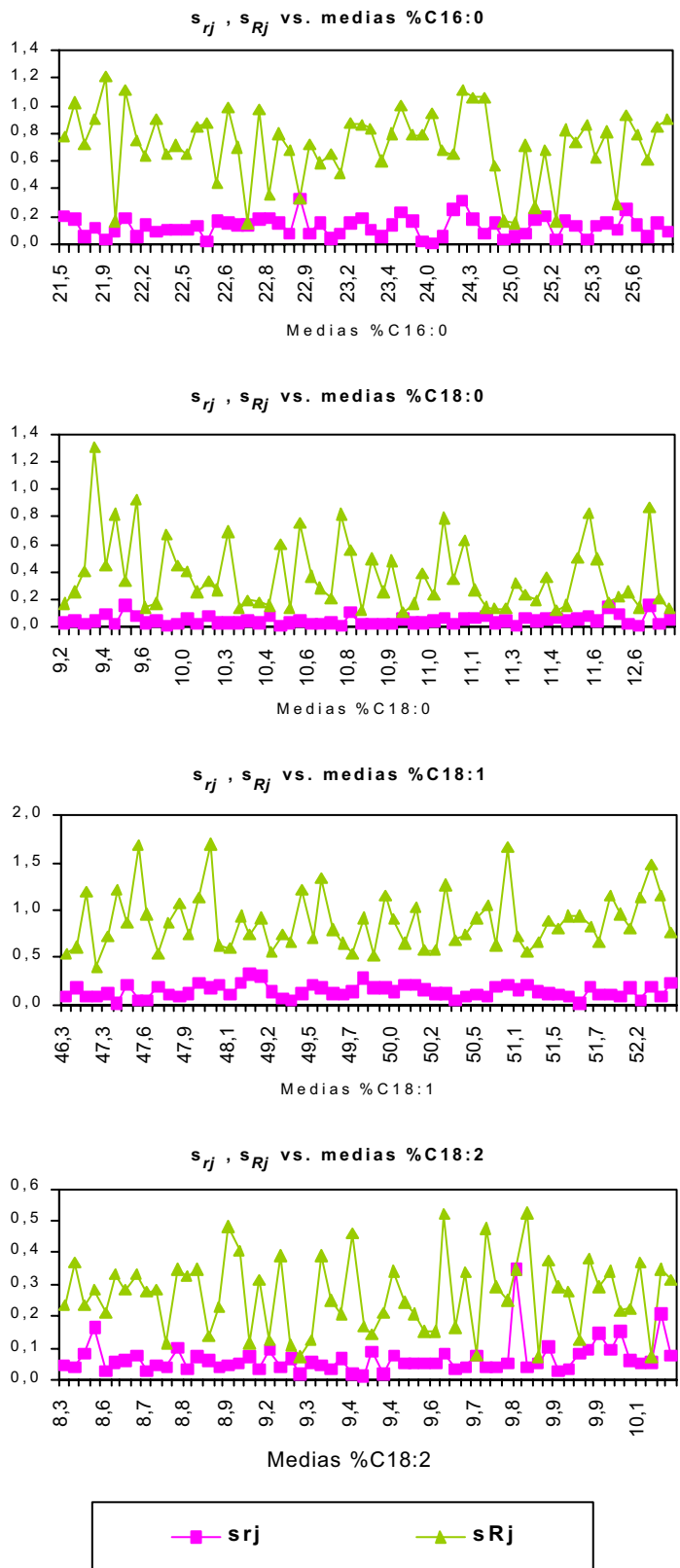
Una vez eliminados dichos datos, se procedió a calcular, para cada una de las muestras, los estadísticos de desviación típica de la repetibilidad (s_{rj}) y desviación típica de la reproducibilidad (s_{Rj}).

Dicho valores oscilan, en general y para todos los ácidos grasos, entre 0 y 0,2 para s_{rj} y entre 0,2 y 1,5 para s_{Rj} . Para cada uno de los ácidos grasos considerados, se han representado los valores de s_{rj} y s_{Rj} , frente a los valores medios, ordenados de forma ascendente, de cada una de las muestras analizadas (Figura 4.1.3). Con ello es posible evaluar la existencia de algún tipo de relación entre dichas desviaciones típicas y los valores medios.

Como puede observarse en la Figura 4.1.3, los estimadores de la precisión no incrementan al aumentan los valores medios de las muestras para todos los ácidos grasos. Por tanto, no parece existir ninguna dependencia entre dichos valores por lo que los estadísticos de la desviación típica de la repetibilidad y reproducibilidad generales (s_r y s_R) pueden ser estimados como la media de dichas desviaciones típicas para todas las muestras.

Dichos resultados, junto con la media global, los valores de repetibilidad y reproducibilidad y las correspondientes desviaciones estándar relativas (RSD_r y RSD_R) aparecen reflejados en la Tabla 4.1.14.

Figura 4.1.3. Representación de s_{rj} y s_{Rj} frente a los valores medios de los ácidos grasos.



Se ha de recordar que los valores de repetibilidad (r) y reproducibilidad (R) informan de la máxima diferencia absoluta (para un 95% de los casos) entre 2 resultados, y solo 2, obtenidos sobre la misma muestra ya sea en un mismo laboratorio (repetibilidad) o en 2 laboratorios diferentes (reproducibilidad).

Tabla 4.1.14. Parámetros estadísticos de repetibilidad y reproducibilidad

Acido Graso	Media Global \bar{y}	DT Repetibilidad s_r	DT Reproducibilidad s_R	Repetibilidad r	Reproducibilidad R	RSD_r (%)	RSD_R (%)
C16:0 (%)	23,60	0,13	0,71	0,37	2,00	0,56	3,04
C18:0 (%)	10,78	0,05	0,38	0,15	1,05	0,49	3,54
C18:1 (%)	49,71	0,15	0,88	0,41	2,47	0,29	1,78
C18:2 (%)	9,37	0,07	0,27	0,18	0,75	0,69	2,86

Como se observa, los valores de repetibilidad (r) oscilan entre 0,15 (para el C18:0) y 0,41 (para C18:1). Como es lógico, los valores de reproducibilidad (R) son sensiblemente más altos y oscilaron entre 0,75 (para el C18:2) y 2,47 (para el C18:1). Con el fin de poder comparar los estadísticos obtenidos para los diferentes ácidos grasos, se han calculado las desviaciones típicas relativas (RSD_r y RSD_R). Los resultados de RSD_r oscilaron entre 0,29% (para el C18:1) y 0,69% (para el C18:2) mientras que para la RSD_R se situaron entre 1,78% (para el C18:1) y 3,54% (para el C18:0).

En la norma 963.22 de la AOAC (1999) se especifica la precisión que ha de tener el método de análisis empleado en este estudio, indicando los valores máximos de repetibilidad y reproducibilidad que han de presentar los ácidos grasos mayoritarios (>5%) en las muestras de grasas o aceites. Así, los valores de r y R han de ser inferiores a 1 y 3 respectivamente, mientras que los valores de RSD_r no deben ser superiores a 3% y los valores de RSD_R no deben superar el 10%. Como se observa en la Tabla 4.1.14, los valores de r , R , RSD_r y RSD_R para los 4 ácidos grasos analizados en este estudio son inferiores a los valores recomendados como máximos por la AOAC por lo que se puede concluir que la precisión obtenida en este estudio ha de ser considerada como adecuada para esta metodología de análisis.

Al mismo tiempo, estos resultados son de mayor precisión a los obtenidos por Firestone and Horwitz (1979) en la determinación de ácidos grasos por cromatografía de gases en grasa de mantequilla (RSD_r C16:0 = 1,88%; RSD_R C16:0 =

3,88%; RSD_r C18:0 = 0,74%; RSD_R C18:0 = 10,52%; RSD_r C18:1 = 1,09%; RSD_R C18:1 = 3,45%; RSD_r C18:2 = 5,45%; RSD_R C18:2 = 13,22%) y del mismo orden a los obtenidos por Cert *et al.* (1996) en aceites de oliva con similar composición de ácidos grasos (RSD_r C16:0 = 1,44-1,78%; RSD_R C16:0 = 3,44-4,45%; RSD_r C18:0 = 0,66-1,97%; RSD_R C18:0 = 4,77-5,46%; RSD_r C18:1 = 0,27-0,49%; RSD_R C18:1 = 0,95-1,17%; RSD_r C18:2 = 0,49-1,94%; RSD_R C18:2 = 2,10-3,29%).

Por tanto, y a modo de conclusión indicar que los resultados de repetibilidad, reproducibilidad, desviación estándar relativa de la repetibilidad y desviación estándar relativa de la reproducibilidad obtenidos para los 4 ácidos grasos mayoritarios en grasa de cerdo Ibérico muestran que la precisión obtenida mediante cromatografía de gases ha de ser considerada como notable y adecuada para este método de análisis así como del mismo orden a la obtenida en aceites de similar composición.

4.2 PUESTA A PUNTO DE METODOLOGÍAS PARA OBTENCIÓN DE INFORMACIÓN ESPECTROSCÓPICA NIR EN GRASA FUNDIDA Y TEJIDO ADIPOSO.

4.2.1 MODALIDAD DE DOBLE TRANSMISIÓN

Un criterio para evaluar la precisión de la tecnología NIRS ha consistido en la medida de la repetibilidad de la señal espectroscópica NIR en grasa fundida de cerdo Ibérico, obtenida mediante la modalidad de doble transmisión. Para ello se utilizó el estadístico *RMS* (apartado 3.2.1).

En este trabajo de optimización de los errores que podían afectar a la señal espectroscópica NIR en grasa fundida, se observó que una posible fuente de error podía ser el estado en que se encontraba la muestra de grasa, ya sea líquida (fundida) o pastosa (estado intermedio entre líquida propiamente dicha y sólida).

Por ello, tras obtener 10 espectros de una misma muestra sin vaciar el contenido de la cápsula, tanto en estado pastoso (24°C) como en estado líquido (40°C), se calcularon los valores de *RMS* entre 30 diferentes combinaciones de pares de espectros, tanto para grasa líquida como pastosa. Los valores medios de los estadísticos *MEAN* y *STD* de *RMS* (apartado 3.2.1), para 30 combinaciones de espectros en estado líquido y 30 combinaciones entre espectros en estado pastoso, aparecen reflejados en la Tabla 4.2.1.

Tabla 4.2.1. Valores medios de estadísticos MEAN y STD entre espectros de una misma cápsula con grasa en estado líquido y pastoso.

Espectros	MEAN	STD
Media estado líquido (40°C) ¹	553	783
Media estado pastoso (24°C) ¹	649	917
Media Pastoso vs. Líquido ¹	2.874	4.064

1: valores medios de 30 combinaciones de pares de espectros

En dicha tabla se observa que los valores medios de *MEAN* y *STD* tanto para muestras en estado líquido (553 y 783 $\mu\log$ 1/R), en estado pastoso (649 y 917 $\mu\log$ 1/R) son bajos y al mismo tiempo similares entre sí lo cual indica la alta repetibilidad de los espectros NIR obtenidos mediante esta modalidad, independientemente del estado en que se encuentre la muestra. Se ha de tener en cuenta que los valores medios de *RMS* para espectros de muestras de productos agrícolas en estado sólido y pulverulento, obtenidos mediante la modalidad de reflexión y el uso de cápsulas selladas, suelen ser inferiores a 200 $\mu\log$ (1/R) por lo que los valores descritos previamente pueden ser considerados bajos teniendo en cuenta que se trata de cápsulas no selladas.

Sin embargo, al realizar las comparaciones entre pares de espectros obtenidos unos en estado líquido y otros en estado pastoso (Tabla 4.2.1), los resultados de *MEAN* y *STD* fueron más elevados (2.874 y 4.064 $\mu\log$ 1/R). Esto denota la influencia de la temperatura de la muestra sobre los espectros obtenidos e indica la idoneidad de mantener un estado único (líquido o pastoso) para el desarrollo de calibraciones con el fin de poder minimizar errores en la toma de espectros.

Estos resultados coinciden con los obtenidos en la Figura 4.2.1 donde han sido representados 10 espectros de una misma cápsula conteniendo una muestra de grasa, tanto en estado líquido como en estado pastoso. Un primer comentario acerca de dicha figura es que en el espectro NIR de grasa fundida de cerdo Ibérico obtenido mediante esta modalidad de análisis, se pueden apreciar nítidamente los picos de absorción característicos de grasas y aceites en la región de 1.700 a 1.800nm, alrededor de 2.150nm y los picos de absorción máxima entre 2.300 y 2.400 nm (apartado 2.3.1).

Un segundo comentario sobre la Figura 4.2.1 es que no existen diferencias apreciables entre los espectros de grasa en estado líquido o pastoso, a excepción del

intervalo comprendido entre 400 y 1.100 nm donde es posible encontrar pequeñas diferencias entre los espectros obtenidos en ambos estados. Al representar con más detalle esta zona del espectro (Figura 4.2.2), no se observan diferencias entre los 10 espectros obtenidos en un mismo estado, corroborando la alta repetibilidad de la información espectral.

Figura 4.2.1. 10 espectros de una cápsula conteniendo muestra en estado líquido y pastoso.

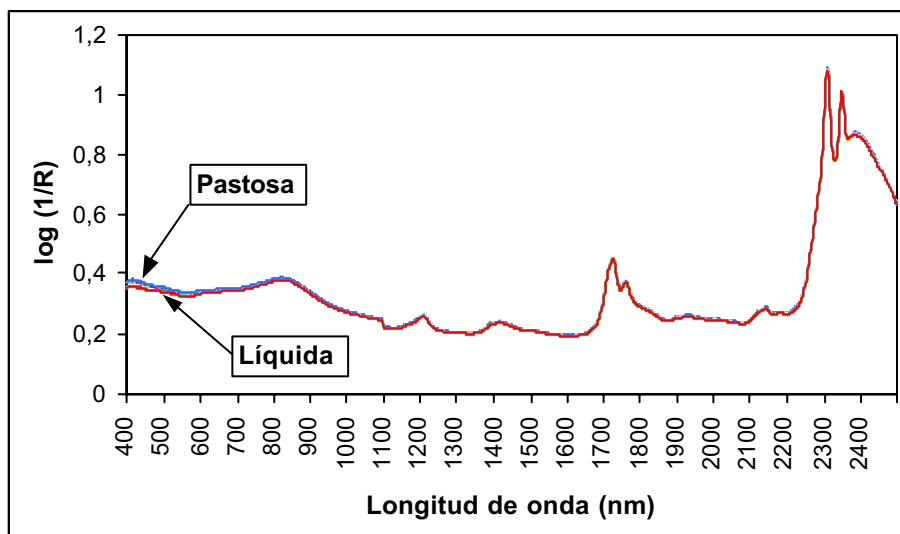
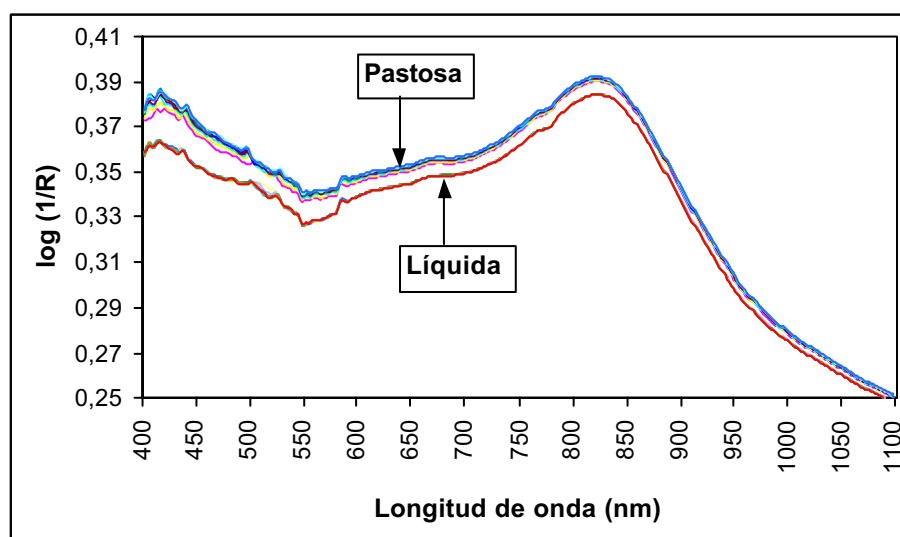


Figura 4.2.2. 10 espectros de 1 cápsula conteniendo muestra en estado líquido y pastoso (detalle entre 400 y 1.100 nm).



Los resultados obtenidos permiten poder decantarse por cualquier estado a la hora de presentar la muestra al instrumento NIR, desde el punto de vista de la repetibilidad de la señal. Dadas las características de la grasa de cerdo Ibérico, la cual

solidifica a diferentes temperaturas (entre el rango de 20°C y 30°C) según el perfil de ácidos grasos debido al efecto del régimen alimenticio a que haya sido sometido el cerdo en la fase de cebo, es difícil poder disponer de las muestras con el mismo grado de solidificación en el momento del análisis NIRS. A una temperatura de 24°C, aquellas muestras procedentes de animales cebados en montanera se encontrarán en estado semilíquido mientras que las muestras procedentes de animales cebados a base de pienso estarán en estado sólido.

Teniendo en cuenta este hecho y los resultados de *RMS* detallados con anterioridad, se recomienda obtener los espectros NIR de las muestras de grasa de cerdo Ibérico en estado líquido, eligiendo para ello una temperatura moderada (ej.: 40°C) en la cual cualquier muestra de grasa de cerdo Ibérico se encontrará en estado líquido, independientemente del tipo de alimentación que haya tenido el animal del que proviene.

Otra magnitud de error también de interés en el cálculo de la repetibilidad espectral es el debido a las diferencias entre submuestras de la misma muestra (efecto de rellenado de la cápsula). Para evaluar dicho error se obtuvieron los espectros de 10 submuestras de una misma muestra en estado líquido, calculándose a continuación los estadísticos *RMS* entre 20 combinaciones de pares de espectros de diferentes submuestras. En la Tabla 4.2.2 se muestran los valores máximos, mínimos y medios de los estadísticos *MEAN* y *STD* calculados.

Tabla 4.2.2. Valores máximos, mínimos y medios de MEAN y STD entre submuestras de la misma muestra.

	MEAN	STD
Valores mínimos	763	1.079
Valores máximos	4.240	5.996
Valores Medios ¹	2.236	3.136

1: valores medios de 20 combinaciones de pares de espectros

Como se observa en dicha tabla, si bien se pueden llegar a obtener valores bajos para los estadísticos *MEAN* y *STD* (763 y 1.079 $\mu\log$ 1/R), similares a los obtenidos entre espectros de una misma cápsula sin rellenar, los valores medios fueron de 2.236 $\mu\log$ (1/R) para el estadístico *MEAN* y 3.136 $\mu\log$ (1/R) para el estadístico

STD. Los valores obtenidos son inferiores a los reseñados por Cobo (1999) en el caso de espectros provenientes de muestras de aceite de oliva.

A partir del valor medio de *STD* y de acuerdo con la fórmula descrita en el apartado 3.2.1.3, el valor del estadístico RMS_{limite} , o valor a partir del cual las submuestras no han de considerarse representativas de la grasa analizada, es 3.249.

En la Figura 4.2.3 se han representado los espectros de 10 submuestras pertenecientes a una misma muestra.

Figura 4.2.3. Espectros de 10 submuestras de la misma muestra.

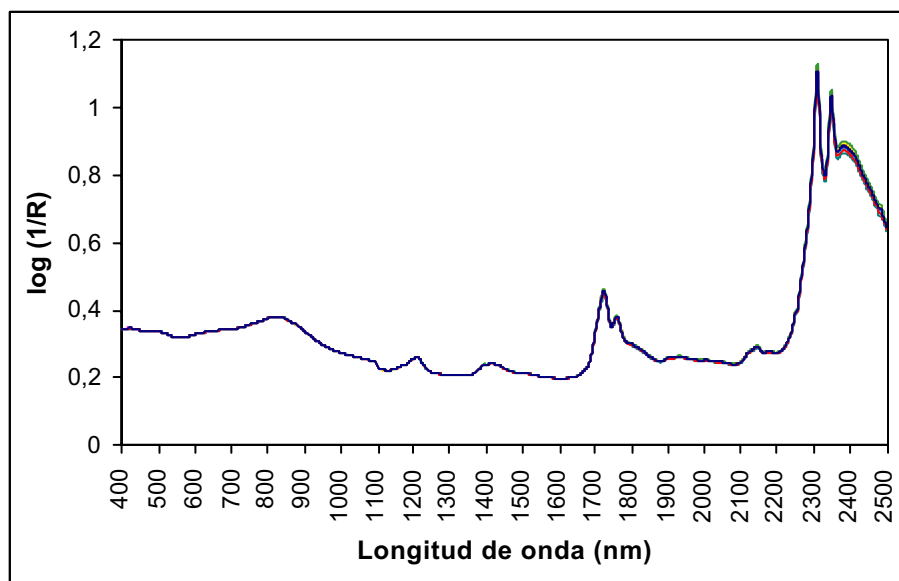
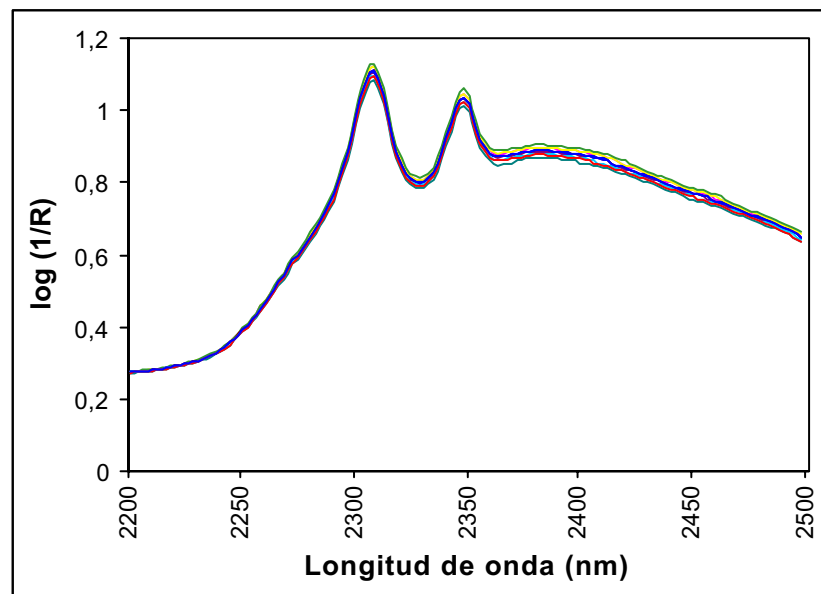


Figura 4.2.4. Espectros de 10 submuestras de la misma muestra.

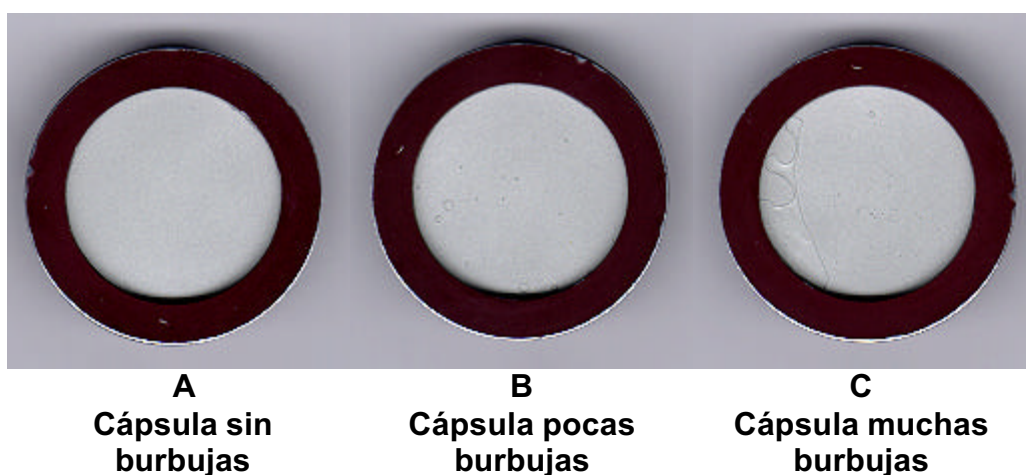


En ella se aprecia que todos los espectros están superpuestos excepto en la región comprendida entre 2.300 y 2.500 nm donde se observan pequeñas diferencias entre ellos (Figura 4.2.4).

Un factor operacional que podría tener gran influencia en la repetibilidad de la medida espectroscópica mediante esta modalidad de análisis NIRS es la presencia de burbujas de aire en el interior de la muestra, tanto en estado líquido como en estado pastoso. Estas burbujas surgen con facilidad al proceder al cierre de la cápsula, siendo mayor la probabilidad de que aparezcan cuanto menos se haya limpiado y secado el cristal y el fondo reflectante de la cápsula.

Con el fin de evaluar el efecto de dicho factor, se provocó deliberadamente la presencia de burbujas en el interior de la cápsula, obteniendo 9 espectros de una misma muestra de grasa líquida de cerdo Ibérico de forma que 3 de ellos correspondían a cápsulas donde no se obtuvieron burbujas en su relleno, 5 a cápsulas con poca presencia de burbujas (entre 10 y 15) y 1 a una cápsula con gran cantidad de aire en su interior (aproximadamente, el 40% del volumen). Las diferencias entre estos tres tipos de cápsulas puede apreciarse en la Figura 4.2.5.

Figura 4.2.5. Cápsulas conteniendo grasa de cerdo con diferente contenido de burbujas de aire.



La Tabla 4.2.3 muestra los valores medios *MEAN* y *STD* para todas las comparaciones entre pares de espectros con diferente presencia de burbujas de aire en su interior. Al comparar espectros obtenidos a partir de cápsulas sin presencia de burbujas de aire en su interior (A *versus* A), los valores resultantes de *MEAN* y *STD* son

similares a los obtenidos en la evaluación de las diferencias entre submuestras de una misma muestra, es decir, del orden de 4.000 (Tabla 4.2.2).

Tabla 4.2.3. Valores Medios de RMS entre espectros de grasa analizada en cápsulas con diferentes proporción de burbujas de aire en su interior.

	MEAN	STD
Media A vs. A*	2.891	4.088
Media A vs. B*	3.223	4.558
Media A vs. C*	25.047	35.420

*A: cápsulas sin burbujas en su interior; B: cápsulas con poca presencia de burbujas de aire; C: cápsulas con gran cantidad de aire en su interior.

Al mismo tiempo, los valores medios de dichos estadísticos obtenidos en la comparación de espectros de cápsulas sin burbujas (A) y aquellos obtenidos en cápsulas con poca presencia de burbujas en su interior (B), son ligeramente superiores si bien se pueden considerar del mismo orden ($STD \cong 4.500$). Estos resultados sugieren que la presencia de un número pequeño de burbujas en el interior de la cápsula a la hora de realizar la toma de espectros, no provoca una diferencia de importancia entre los espectros obtenidos.

Sin embargo, cuando se comparan espectros de cápsula sin burbujas de aire (A) con cápsulas que contengan una gran cantidad de burbujas en su interior (C), los valores medios de *MEAN* y *STD* son muy elevados lo cual nos indica la gran diferencia existente entre ambos tipos de espectros. Estas diferencias pueden apreciarse en la Figura 4.2.6 y de forma más nítida en la Figura 4.2.7.

Como se observa en dichas figuras, existe una clara diferencia entre espectros obtenidos en cápsulas sin burbujas o con poca cantidad de burbujas de aire y espectros de cápsulas con una elevada proporción de burbujas de aire en su interior, fundamentalmente en la región comprendida entre 2.200 y 2.500 nm. En ella, los espectros de cápsulas con gran cantidad de burbujas presentan valores de absorbancia inferiores, probablemente debidos a la menor cantidad de muestra de grasa existente en la cápsula.

Figura 4.2.6. Espectros de cápsulas con presencia de burbujas de aire.

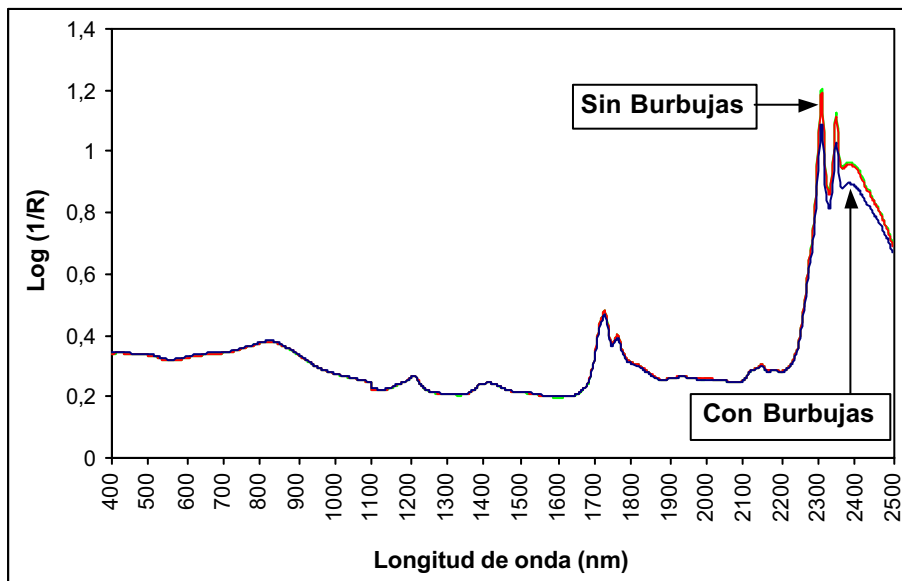
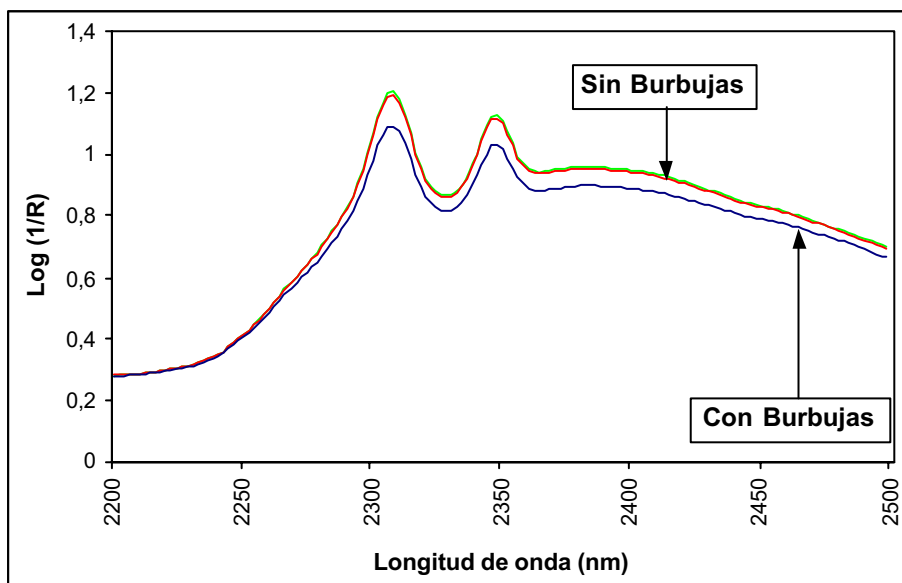


Figura 4.2.7. Detalle de espectros de cápsulas con presencia de burbujas de aire.



Por todo lo expuesto anteriormente, la metodología de análisis de este tipo de grasa mediante la modalidad de doble transmisión consistiría en la obtención de espectros NIR por duplicado, considerando un valor de RMS_{limite} aproximado de 3.500 a 4.000 $\mu\log(1/R)$, manteniendo la muestra a 40°C mediante ventilación forzada, baño termostático o placa calefactora y evitando la presencia de burbujas a la hora del llenado de la cápsula.

4.2.2 MODALIDAD DE INTERACTANCIA-REFLECTANCIA

Como se ha indicado en el apartado 3.2.2, en este punto se trataba de evaluar el efecto de diversos factores sobre la repetibilidad de los datos espectrales obtenidos mediante la modalidad de análisis de interactancia-reflectancia. Estos factores fueron el número de espectros a tomar de una muestra, la localización de la sonda en una muestra de tejido adiposo y el efecto de luz ambiental.

Para evaluar el número de espectros a tomar y la localización de la sonda en la muestra, se seleccionó una muestra de grasa subcutánea en la que cual se apreciaran las 2 dos capas de tejido adiposo (externa e interna). En dicha muestra se obtuvieron espectros en 8 posiciones, 4 de ellas situadas en la capa externa y 4 en la capa interna de tejido adiposo.

A su vez, para evaluar la repetibilidad de la señal espectroscópica se obtuvieron 4 espectros sobre cada una de las localizaciones sin mover la sonda de fibra óptica. A continuación se calcularon los valores de *RMS*, *MEAN* y *STD* entre diferentes combinaciones de grupos de 2 espectros, evaluando el error existente dentro de cada posición de análisis, el error entre espectros tomados en diferentes localizaciones de la misma capa grasa y el error entre espectros de diferentes capas de grasa subcutánea.

En la Tabla 4.2.4 se muestran los valores de los estadísticos *MEAN* y *STD* correspondientes a los 4 espectros obtenidos, sin mover la sonda de fibra óptica, para cada una de las localizaciones.

Tabla 4.2.4. Valores de RMS en una misma posición de la sonda en tejido adiposo.

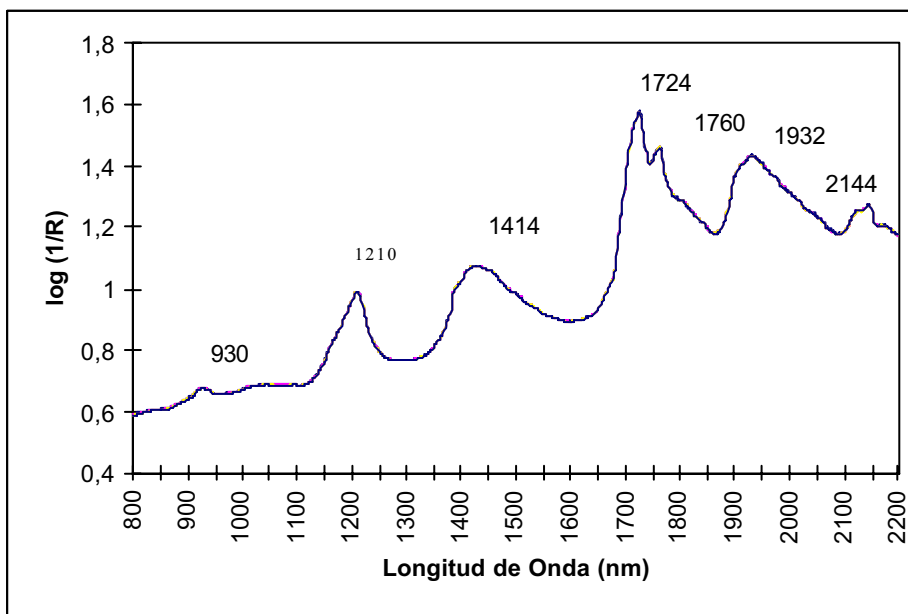
	MEAN	STD		MEAN	STD
Loc E1	1.145	1.322	Loc I1	4.679	5.403
Loc E2	869	1.004	Loc I2	2.489	2.874
Loc E3	2.732	3.055	Loc I3	1.323	1.528
Loc E4	6.864	7.926	Loc I4	2.033	2.348
Media	2.903	3.327	Media	2.631	3.038

Como puede observarse, los valores de *STD* oscilaron desde un mínimo de 1.004 $\mu\log(1/R)$ para la posición Loc E2 a un máximo de 7.926 $\mu\log(1/R)$ para la posición Loc E4. Sin embargo, se obtuvieron valores de *STD* medios muy similares

tanto para los espectros tomados en las posiciones de la capa interna ($3.038 \mu\log 1/R$) como los de la externa ($3.327 \mu\log 1/R$).

Esta alta repetibilidad de los valores espectrales se puede observar de forma muy clara al representar los espectros obtenidos en una misma posición. A modo de ejemplo, la Figura 4.2.8 muestra 4 espectros obtenidos en la posición Loc I3 sin mover la sonda de fibra óptica de su localización en la muestra de tejido adiposo. En ella se observa la práctica superposición de los 4 espectros obtenidos y por tanto la alta repetibilidad de la señal espectral NIR dentro de una misma posición en la muestra.

Figura 4.2.8. 4 espectros en la posición Loc I3.



Se ha de destacar en dicha figura que los espectros de tejido adiposo presentan las bandas de absorción NIR esperadas de grasa (1.210, 1.724, 1.760 y 2.144 nm) junto a otras debidas probablemente a la presencia de tejido conectivo y agua en el tejido adiposo, localizadas en regiones de absorción de los enlaces $-NH$ y $-OH$ (1.414 y 1.932 nm).

Para el cálculo de diferencias existentes entre espectros obtenidos en diferentes posiciones de la fibra óptica sobre la muestra, se obtuvieron los espectros medios para cada una de las localizaciones a partir de los 4 espectros tomados inicialmente. Mediante las comparaciones 2 a 2 de dichos espectros medios, se calcularon los valores *MEAN* y *STD* de *RMS*, mostrándose los resultados en las Tablas 4.2.5 a 4.2.7.

Tabla 4.2.5. Valores de RMS entre diferentes posiciones, en capa externa, de la sonda en tejido adiposo.

	MEAN	STD
Loc E1 vs. Loc E2	29.422	41.609
Loc E1 vs. Loc E3	54.435	76.980
Loc E1 vs. Loc E4	24.725	34.966
Loc E2 vs. Loc E3	27.469	38.846
Loc E2 vs. Loc E4	10.777	15.241
Loc E3 vs. Loc E4	31.132	44.027
Media de 6 comparaciones	29.660	41.945

Tabla 4.2.6. Valores de RMS entre diferentes posiciones, en capa interna, de la sonda en la muestra.

	MEAN	STD
Loc I1 vs. Loc I2	46.848	66.251
Loc I1 vs. Loc I3	79.771	112.806
Loc I1 vs. Loc I4	52.573	74.347
Loc I2 vs. Loc I3	33.250	47.021
Loc I2 vs. Loc I4	10.476	14.815
Loc I3 vs. Loc I4	28.569	40.402
Media de 6 comparaciones	41.915	59.274

Tabla 4.2.7. Valores de RMS entre posiciones de la sonda en diferentes capas de tejido adiposo (Externa vs. Interna).

	MEAN	STD
Loc E1 vs. Loc I1	26.625	37.653
Loc E1 vs. Loc I2	31.802	44.974
Loc E1 vs. Loc I3	61.416	86.852
Loc E1 vs. Loc I4	34.575	48.896
Loc E2 vs. Loc I1	52.487	74.225
Loc E2 vs. Loc I2	17.993	25.445
Loc E2 vs. Loc I3	34.215	48.386
Loc E2 vs. Loc I4	15.929	22.526
Loc E3 vs. Loc I1	72.954	103.167
Loc E3 vs. Loc I2	28.981	40.985
Loc E3 vs. Loc I3	17.418	24.632
Loc E3 vs. Loc I4	28.315	40.043
Loc E4 vs. Loc I1	46.520	65.787
Loc E4 vs. Loc I2	19.865	28.093
Loc E4 vs. Loc I3	41.113	58.141
Loc E4 vs. Loc I4	21.925	31.006
Media de 16 comparaciones	34.508	48.801

Como se observa en dichas tablas, los resultados de *RMS* obtenidos en el cálculo de las diferencias espectrales entre diferentes posiciones de muestreo (41.945 $\mu\log$ 1/R, 59.274 $\mu\log$ 1/R y 48.801 $\mu\log$ 1/R) son muy superiores a los valores

obtenidos en una misma posición (3.038 $\mu\log$ 1/R y 3.327 $\mu\log$ 1/R), debido a las lógicas diferencias existentes en el material heterogéneo a analizar como es el tejido adiposo de cerdo Ibérico.

Además, se puede apreciar en la Tabla 4.2.5 que los valores de *MEAN* y *STD* obtenidos tras las comparaciones de los espectros medios de las localizaciones de la capa externa (29.660 y 41.945 $\mu\log$ 1/R) son del mismo orden de magnitud que los valores obtenidos (Tabla 4.2.6) para las comparaciones de los espectros de la capa interna (41.915 y 59.274 $\mu\log$ 1/R) y a su vez semejantes a los valores resultantes (Tabla 4.2.7) de las comparaciones de espectros obtenidos en diferentes capas (34.508 y 48.801 $\mu\log$ 1/R).

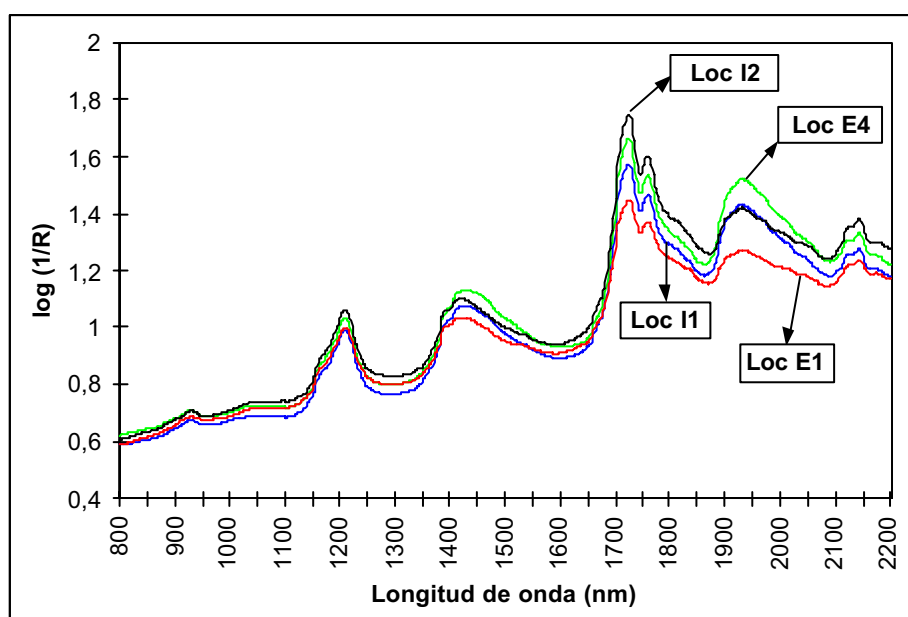
A raíz de estos resultados, se puede concluir que para obtener un espectro representativo de una muestra de tejido adiposo y dada la alta repetibilidad de las medidas espectrales obtenidas en posiciones de la misma capa o de diferentes capas de la muestra, la toma de espectros NIR mediante la modalidad empleada se puede realizar, independientemente en cualquiera de las capas del tejido adiposo. Sin embargo, dadas las diferencias detectadas en cuanto a composición de ácidos grasos en las diferentes capas del tejido adiposo de cerdo Ibérico por Casillas (1995), sería preferible obtener espectros en localizaciones de diferentes capas, siempre y cuando estas se aprecien en las muestras.

Tras observar de forma más detallada los resultados obtenidos de la comparación de pares de espectros, puede apreciarse que los valores de *MEAN* y *STD* en aquellos pares donde interviene la posición Loc I1 son, en general, superiores a los obtenidos en otros pares. Estas mayores diferencias espectrales al intervenir en la comparación la posición Loc I1 se pueden deber a una mayor diferencia del espectro medio de dicha localización frente al resto de espectros medios.

En la Figura 4.2.9 se muestran los espectros medios para las posiciones Loc I1, Loc I2, Loc E1 y Loc E4. En ella se aprecia que los valores de absorbancia para el espectro obtenido en la posición Loc I1 son inferiores a los valores espectrales obtenidos en otras localizaciones, fundamentalmente en el rango comprendido entre 1.700 y 2.200 nm. Durante la toma de espectros en la posición Loc I1, se observó que la ubicación de la sonda respecto a la muestra no era completamente perpendicular sino

que existía una ligera desviación, provocando una ausencia de contacto pleno entre la ventana de la sonda y la superficie de la muestra. Esto pudo provocar la presentación de una menor cantidad de muestra al instrumento y, por tanto, una disminución de los valores de absorbancia. Por tanto, y dada la importancia que podía suponer este factor, se tuvo especial precaución en lograr un contacto pleno entre la sonda y la muestra, a ser posible, en posición perpendicular.

Figura 4.2.9. 4 espectros en 4 posiciones diferentes de la muestra.



Otro factor operacional que podría influir en el uso de una sonda de fibra óptica de las características de la empleada en este trabajo es la presencia luz externa que pudiera alcanzar los detectores del instrumento a través de la sonda. Esta presencia de luz podría deberse a la ausencia de un contacto imperfecto entre la sonda y la superficie de la muestra o bien a la penetración de la radiación exterior a través de la muestra.

Para evaluar este efecto sobre la repetibilidad espectral, se obtuvieron 4 espectros en 3 diferentes posiciones de una muestra (Loc A, Loc B y Loc C) tanto protegiendo como sin proteger a la muestra mediante una caja negra que evitara la presencia de luz procedente del exterior.

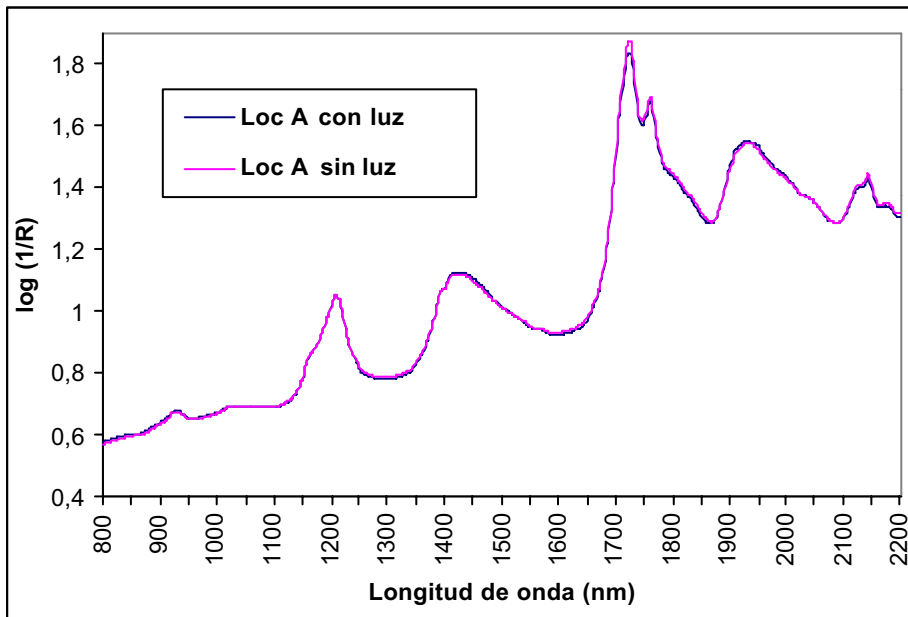
La Tabla 4.2.8 muestra los resultados de los valores de los estadísticos *MEAN* y *STD* al comparar los 4 espectros obtenidos en cada una de las 3 localizaciones mientras que la Figura 4.2.10 muestra el espectro medio para la posición Loc A tanto en

presencia como con ausencia de luz ambiental. En ambos casos se puede apreciar la nula influencia de dicho factor sobre la repetibilidad espectral. Además, los valores de *STD* obtenidos en todas las localizaciones se situaron dentro del rango obtenido con anterioridad (Tabla 4.2.4).

Tabla 4.2.8. Influencia de la luz ambiental en los valores de RMS.

	CON LUZ		SIN LUZ	
	MEAN	STD	MEAN	STD
Loc A	1.865	2.145	1.784	1.995
Loc B	3.172	3.663	1.907	2.133
Loc C	2.327	2.687	1.995	2.230
Media	2.455	2.832	1.895	2.119

Figura 4.2.10. Espectros medios en la misma posición con y sin protección de la luz ambiental.



Los resultados obtenidos revelan que se pueden obtener fácilmente valores de *STD* para el estadístico *RMS* inferiores a 4.000 $\mu\log(1/R)$ entre espectros obtenidos en una misma posición de las muestras de tejido adiposo. Sin embargo, el número de posiciones de medida espectral empleadas para obtener un espectro medio de una muestra de tejido adiposo es crítico. Se considera que, para obtener dicho espectro medio, únicamente es necesario obtener 2 espectros en cada posición de la sonda sobre la muestra si bien se recomienda posicionar la sonda, al menos en 2 localizaciones diferentes sobre la superficie de medida. Además se recomienda emplear un valor límite

de *RMS* de 45.000 $\mu\log$ (1/R) entre espectros obtenidos en diferentes puntos de la muestra.

Durante el desarrollo de aplicaciones NIRS cuantitativas y cualitativas y para poder minimizar las fuentes de error a la hora de obtener un espectro representativo de una muestra, se recomienda tener especial precaución en asegurar un ángulo de 90° y un contacto perfecto entre la sonda y la superficie de muestra de tejido adiposo.

4.2.3 INFLUENCIA DE LA CONGELACIÓN DE MUESTRAS DE GRASA DE CERDO IBÉRICO SOBRE SU INFORMACIÓN ESPECTRAL NIR

El diseño experimental de este estudio, descrito en el apartado 3.2.3, se realizó para poder evaluar la influencia de la congelación sobre la medida espectroscópica NIR de la grasa fundida de cerdo Ibérico. Para ello se estudió la influencia tanto del tiempo de conservación de las muestras bajo congelación y el efecto de sucesivas descongelaciones y congelaciones de la muestra para su análisis NIRS.

Para estudiar la influencia de la conservación de las muestras bajo congelación, se utilizaron espectros de muestras de grasa que permanecieron congeladas diferentes periodos de tiempo y que no se habían visto afectadas por el efecto de la congelación-descongelación. Es decir, de acuerdo con el diseño experimental descrito en el apartado 3.2.3.2, se seleccionaron aquellos espectros pertenecientes a muestras cuyos espectros fueron tomados a la apertura por vez primera de los tubos de muestras ($T_{0,0}$ a $T_{5,0}$).

En el análisis de este factor, todas las comparaciones entre las diferentes poblaciones espectrales obtenidas en cada mes de estudio se llevaron a cabo estableciendo como testigo aquellos espectros obtenidos en el primer mes de análisis NIRS. Matizar que dicho análisis se realizó en la misma fecha en la que se obtuvieron las muestras de grasa líquida del tejido adiposo.

Con se ha indicado en el apartado 3.2.3.3, se han empleado dos estrategias diferentes para la evaluación de diferencias espectrales. Los resultados de la aplicación de la primera de ellas, esto es, el cálculo del estadístico *RMS(C)*, permite obtener una

estimación de la repetibilidad espectral a partir de las variables originales (valores de absorbancia a diferentes longitudes de onda).

A fin de poder obtener una estimación de la repetibilidad espectral en ausencia del tratamiento de congelación, se compararon 60 espectros pertenecientes a 10 submuestras con diferentes tiempo de conservación bajo congelación. Los valores de $RMS(C)$ para el rango visible y NIR de dichas comparaciones se muestran en la Tabla 4.2.9. Asimismo, en la citada tabla se indican los valores de dichos estadísticos para comparaciones de espectros obtenidos a partir de submuestras de las mismas muestras sin someterse a congelación. En este caso, los valores obtenidos oscilan entre 138 y 304 $\mu\log(1/R)$ para $RMS(C)$ del Visible y entre 3.131 y 5.953 $\mu\log(1/R)$ para $RMS(C)$ del rango NIR.

Tabla 4.2.9. $RMS(C)$ en el rango Visible y NIR entre poblaciones espectrales con diferente tiempo de conservación bajo congelación.

Comparación	$RMS(C)$ VIS	$RMS(C)$ NIR
$T_{0,0a}$ vs. $T_{0,0b}$	138	5.953
$T_{0,0c}$ vs. $T_{0,0d}$	304	3.131
$T_{1,0}$ vs. $T_{0,0}$	1.423	6.133
$T_{2,0}$ vs. $T_{0,0}$	1.856	5.861
$T_{3,0}$ vs. $T_{0,0}$	2.118	7.320
$T_{4,0}$ vs. $T_{0,0}$	2.665	3.673
$T_{5,0}$ vs. $T_{0,0}$	3.624	4.102

Como se aprecia en la Tabla 4.2.9, los valores de $RMS(C)$ del Visible para las comparaciones de espectros obtenidos tras varios meses de congelación (1.423 a 3.624 $\mu\log(1/R)$) resultaron ser muy superiores a los obtenidos entre espectros de submuestras sin congelar (138 y 304 $\mu\log(1/R)$). Al mismo tiempo, se observa que los valores de $RMS(C)$ del Visible se incrementan conforme mayor es el tiempo de conservación bajo congelación. Así, para la comparación de los espectros tomados en el Mes 1 ($T_{1,0}$) frente a los obtenidos en el Mes 0 ($T_{0,0}$), el valor de $RMS(C)$ fue de 1.423 $\mu\log(1/R)$ mientras que para la comparación de espectros tomados en el Mes 5 ($T_{5,0}$) frente a los obtenidos en el Mes 0 ($T_{0,0}$) fue de 3.624 $\mu\log(1/R)$. Es decir, las diferencias encontradas para el estadístico $RMS(C)$ del Visible fueron superiores conforme mayor fue el periodo de congelación considerado.

Los valores de $RMS(C)$ del rango NIR para las comparaciones de espectros obtenidos tras varios meses de congelación oscilaron entre un mínimo de 3.673 y un

máximo de $7.320 \mu\log(1/R)$, valores estos del mismo orden o ligeramente superiores que los obtenidos entre espectros de submuestras sin congelar (entre 3.131 y $5.953 \mu\log(1/R)$). En este caso, no se apreciaron mayores valores de dicho estadístico al incrementarse el tiempo de conservación bajo congelación. Ello mostró las escasas o nulas diferencias existentes en el rango NIR entre espectros de submuestras conservadas bajo congelación durante diferentes periodos de tiempo.

Una segunda estrategia utilizada para la evaluación de la influencia de la congelación sobre la señal espectroscópica fue el cálculo de la distancia H mediante la aplicación del algoritmo CENTER (apartado 3.2.3.3).

En la mayoría de los trabajos sobre la técnica NIRS, la aplicación más empleada del programa CENTER consiste en generar un nuevo espacio vectorial con las componentes principales como variables y calcular, a partir de ahí, la distancia H de cada una de las muestras (espectros) al centro (centroide) de la población o a una muestra determinada. Con ello es posible evaluar la existencia de muestras anómalas al resto de la población.

Existe otra posibilidad de aplicación menos empleada pero de gran utilidad, particularmente para este trabajo, basada en la evaluación de las diferencias entre 2 poblaciones espectrales. Consiste en desarrollar, a partir de un conjunto de espectros A, un espacio vectorial n dimensional (siendo n el número de componentes principales empleadas) y proyectar una población espectral B sobre el espacio vectorial definido por la población A. Seguidamente, es posible calcular la distancia H de cada uno de los espectros y del centroide de la población B al centroide de la población A. Cuanto menor sea la distancia entre ambos centroides, más parecidas serán ambas poblaciones en el espacio n dimensional.

En este estudio se calcularon, mediante esta última modalidad, diferentes estadísticos H dependiendo del rango de variables originales empleadas en el cálculo de las componentes principales. Así, se evaluaron las diferencias espectrales mediante el estudio del estadístico H para la región del Visible (H_{VIS}) y NIR (H_{NIR}), en el espacio vectorial definido por las componentes principales de las variables originales.

En la Tabla 4.2.10 se indican los valores H_{VIS} y H_{NIR} para comparaciones de espectros pertenecientes a muestras con diferentes tiempos de conservación bajo

congelación. Asimismo, en la citada tabla se indican los valores de dichos estadísticos para comparaciones de espectros obtenidos a partir de submuestras de las mismas muestras sin someterse al tratamiento de congelación, siendo en este caso de 0,90 para H_{VIS} y 1,09 para H_{NIR} . Estos valores son del mismo orden independientemente del rango de medida considerado, y al mismo tiempo, inferiores al valor recomendado por Shenk y Westerhaus (1995b) para considerar dos muestras (en este caso, dos poblaciones espectrales) como pertenecientes a una misma población ($H < 3$).

Tabla 4.2.10. H_{VIS} y H_{NIR} entre poblaciones espectrales con diferente tiempo de conservación bajo congelación.

Comparación	H_{VIS}	H_{NIR}
$T_{0,0m}$ vs. $T_{0,0n}$	0,90	1,09
$T_{1,0}$ vs. $T_{0,0}$	5,65	0,80
$T_{2,0}$ vs. $T_{0,0}$	17,36	0,92
$T_{3,0}$ vs. $T_{0,0}$	27,17	1,62
$T_{4,0}$ vs. $T_{0,0}$	74,18	0,75
$T_{5,0}$ vs. $T_{0,0}$	150,08	0,82

Como se puede observar en dicha tabla, los valores de H_{VIS} para las comparaciones de espectros obtenidos tras varios meses de congelación resultaron ser muy superiores a los obtenidos entre espectros de submuestras sin congelar (0,90), incrementándose dichos valores conforme los espectros comparados fueron analizados en meses más distantes. Así, para la comparación de los espectros tomados tras la conservación de las muestras 1 mes de bajo congelación ($T_{1,0}$) frente a los obtenidos sin tratamiento de congelación ($T_{0,0}$), el valor de H_{VIS} fue de 5,65 mientras que para la comparación de espectros tomados en el 5º mes ($T_{5,0}$) frente a los obtenidos inicialmente ($T_{0,0}$) alcanzó valores de 150,08. Por ello, y tras aplicar el límite recomendado por Shenk y Westerhaus (1995b) para considerar una muestra como anómala al resto de la población ($H > 3$), se deberían de considerar diferentes los rangos del Visible de espectros de muestras congeladas durante varios meses a los de las mismas muestras sin congelar.

Los valores de H_{NIR} obtenidos en las comparaciones de espectros de muestras con diferentes periodos de congelación oscilan entre valores de 0,75 y 1,62, siendo estos similares o incluso inferiores a los obtenidos en la comparación de submuestras sin tratamiento de congelación ($H_{NIR}=1,09$). Al mismo tiempo, los valores de H_{NIR} entre los centroides de las poblaciones para las diferentes comparaciones entre

espectros son inferiores a 3 y similares a 0,6 por lo que, de acuerdo con los criterios de Shenk y Westerhaus (1995b), se pueden llegar a considerar dichas poblaciones como vecinas, considerando únicamente el rango NIR de los espectros.

Por tanto, la conservación de muestras de grasa bajo congelación durante periodos prolongados provoca claras diferencias espectrales en la región del Visible (400 a 1100 nm) debidas posiblemente a cambios de coloración provocados por fenómenos de oxidación, presumiblemente iniciados durante la etapa de fusión del tejido adiposo mediante microondas (Yoshida, 1993). Estos fenómenos de oxidación, una vez iniciados, podrían ser ralentizados a bajas temperaturas pero no podrían ser detenidos totalmente (Gray y Pearson, 1987).

Tras un análisis de la información espectral obtenida, es posible detectar las leves alteraciones (cambios de color) que se hayan producido en la región del Visible, aunque estas no sean apreciables a simple vista. Sin embargo, estas diferencias espectrales no se observan en la región NIR, lo cual indica que las alteraciones que puedan provocarse tras una congelación de las muestras, no afecta a la información espectral obtenida en dicha región.

Para estudiar la influencia del efecto de descongelaciones y congelaciones sucesivas sobre la repetibilidad espectral NIR de muestras de grasa, se seleccionaron espectros de muestras con diferentes tratamientos de congelación-descongelación y que hubieran estado conservadas bajo congelación durante un mismo periodo de tiempo para eliminar la influencia de dicho factor. Para ello se compararon 60 espectros obtenidos de 10 submuestras, analizadas en el 5º mes y sometidas a un diferente número de procesos de congelación y descongelación. De acuerdo con el diseño experimental descrito en el apartado 3.2.3.2, dichas muestras corresponderían a las series $T_{5,0}$ a $T_{0,5}$.

La Tabla 4.2.11 muestra los resultados de $RMS(C)$ para los rangos Visible y NIR obtenidos tras la comparación de poblaciones espectrales con diferentes tratamientos de congelación-descongelación. Como puede apreciarse, los valores de $RMS(C)$ para el visible oscilan entre 180 y 241 $\mu\log(1/R)$ mientras que para $RMS(C)$ varían entre 2.633 y 5.685 $\mu\log(1/R)$. En ambos casos, todos los valores de las comparaciones espectrales realizadas, son similares e incluso inferiores a aquellos obtenidos en la comparación de submuestras sin tratamiento de congelación-

descongelación. Esto indica que no se aprecian diferencias espectrales entre muestras sometidas a sucesivos procesos de congelación-descongelación por lo que dicho factor presenta una escasa o nula influencia sobre la información espectral obtenida tanto en la región del Visible como NIR.

Tabla 4.2.11. RMS(C) en el rango Visible y NIR entre poblaciones espectrales con diferentes tratamientos de congelación – descongelación.

Comparación	RMS(C) VIS	RMS(C) NIR
T _{5,0a} vs. T _{5,0b}	138	5.953
T _{5,0c} vs. T _{5,0d}	304	3.131
T _{4,1} vs. T _{5,0}	237	5.685
T _{3,2} vs. T _{5,0}	180	2.633
T _{2,3} vs. T _{5,0}	196	4.338
T _{1,4} vs. T _{5,0}	210	5.193
T _{0,5} vs. T _{5,0}	241	3.620

En la Tabla 4.2.12 se indican los valores H_{VIS} y H_{NIR} para comparaciones de espectros pertenecientes a muestras con diferentes tratamientos de congelación-descongelación. Como se observa, los valores de H_{VIS} y H_{NIR} son similares independientemente del tratamiento a que hayan sido sometidas las muestras, oscilando entre 0,67 y 2,25 para H_{VIS} y entre 1,10 y 1,85 para H_{NIR} . Al mismo tiempo, dichos valores son similares o ligeramente superiores a los obtenidos tras la comparación de espectros pertenecientes a submuestras no sometidas al efecto de la congelación-descongelación ($H_{VIS}=0,90$; $H_{NIR}=1,09$) y en todos los casos inferiores al límite recomendado por Shenk y Westerhaus (1995b) para considerar dos muestras como pertenecientes a poblaciones diferentes ($H>3$).

Tabla 4.2.12. H_{VIS} y H_{NIR} entre poblaciones espectrales con diferente tratamiento de congelación – descongelación.

Comparación	H_{VIS}	H_{NIR}
T _{5,0m} vs. T _{5,0n}	0,90	1,09
T _{4,1} vs. T _{5,0}	2,25	1,59
T _{3,2} vs. T _{5,0}	0,67	1,12
T _{2,3} vs. T _{5,0}	2,24	1,10
T _{1,4} vs. T _{5,0}	2,20	1,85
T _{0,5} vs. T _{5,0}	1,21	1,35

Por tanto, todas las poblaciones espectrales comparadas pueden considerarse como semejantes, por lo que se observa de nuevo la escasa influencia del tratamiento de

congelación-descongelación, en este caso sobre la información espectral definida en el espacio vectorial con las componentes principales como variables.

Los resultados de este estudio son de una enorme repercusión práctica, desde el punto de vista de la implantación de la tecnología NIRS a nivel de laboratorios de control de grasas al ser dichos productos perecederos y fácilmente oxidables.

4.3 DESARROLLO Y EVALUACIÓN DE ECUACIONES DE CALIBRACIÓN NIRS PARA LA PREDICCIÓN DEL CONTENIDO EN ÁCIDOS GRASOS

4.3.1 MODALIDAD DE DOBLE TRANSMISIÓN

4.3.1.1 Obtención de ecuaciones preliminares para la determinación del contenido en ácidos grasos en grasa fundida

Es importante recordar que, de acuerdo con el plan de trabajo propuesto en esta Tesis Doctoral y teniendo en cuenta el compromiso de mostrar al sector del cerdo Ibérico la viabilidad de la técnica NIRS para la predicción del contenido en ácidos grasos en grasa subcutánea, fue necesario el desarrollo de unas ecuaciones preliminares en las primeras etapas del proyecto de investigación.

En concreto, se llegó a un compromiso con la D.O. “Jamón de Huelva” de generar dichas ecuaciones preliminares con las que predecir el contenido en ácidos grasos de muestras de grasa las que, con posterioridad, se les realizarían el análisis de referencia (cromatografía de gases). Con ello la propia Denominación de Origen dispondría de una estimación de la composición media en ácidos grasos en ácidos grasos de cada lote con varios días de antelación a la composición determinada mediante cromatografía de gases y así poder apreciar todas las ventajas de la implantación de la tecnología NIRS. En este apartado se describirá el proceso seguido para el desarrollo de estas ecuaciones preliminares.

Por todo ello, una vez puesta a punto la metodología de obtención de espectros se inició el trabajo de desarrollo de ecuaciones de calibración NIRS, mediante regresión por mínimos cuadrados parciales, para la predicción del contenido en ácidos grasos de grasa fundida de cerdo Ibérico, empleando para ello la modalidad de análisis de doble transmisión.

A partir de los datos de referencia (cromatografía gaseosa) y de los datos espectroscópicos correspondientes a muestras controladas durante las campañas de sacrificio 96/97 y 97/98, se trataba de obtener ecuaciones de calibración para cada uno de los ácidos grasos mayoritarios (palmítico, C16:0; esteárico, C18:0; oleico, C18:1 y linoléico, C18:2) y que son los utilizados en la clasificación de partidas mediante aplicación del contrato tipo homologado. Un paso crítico en el desarrollo de ecuaciones de calibración NIRS es el de la estructura del colectivo de calibración. Dicho colectivo debe recoger todas las fuentes de variación que cabe esperar contendrían las muestras futuras, sobre las que se aplicará la ecuación obtenida y existen diferentes estrategias para la selección de dicho colectivo (Shenk *et al.*, 1992; Shenk y Westerhaus, 1996).

En este caso y teniendo en cuenta que no existía limitante de poseer un número elevado de datos de referencia, lo cual no es habitual en trabajos de desarrollo de calibraciones NIRS, se procedió a realizar una estructura basada en el conocimiento a priori de los datos químicos de referencia. Ello obviamente conlleva el incluir en dicho colectivo partidas de diferente composición en los cuatro ácidos grasos y por tanto provenientes de diferentes regímenes alimenticios (*bellota*, *recebo* y *cebo*).

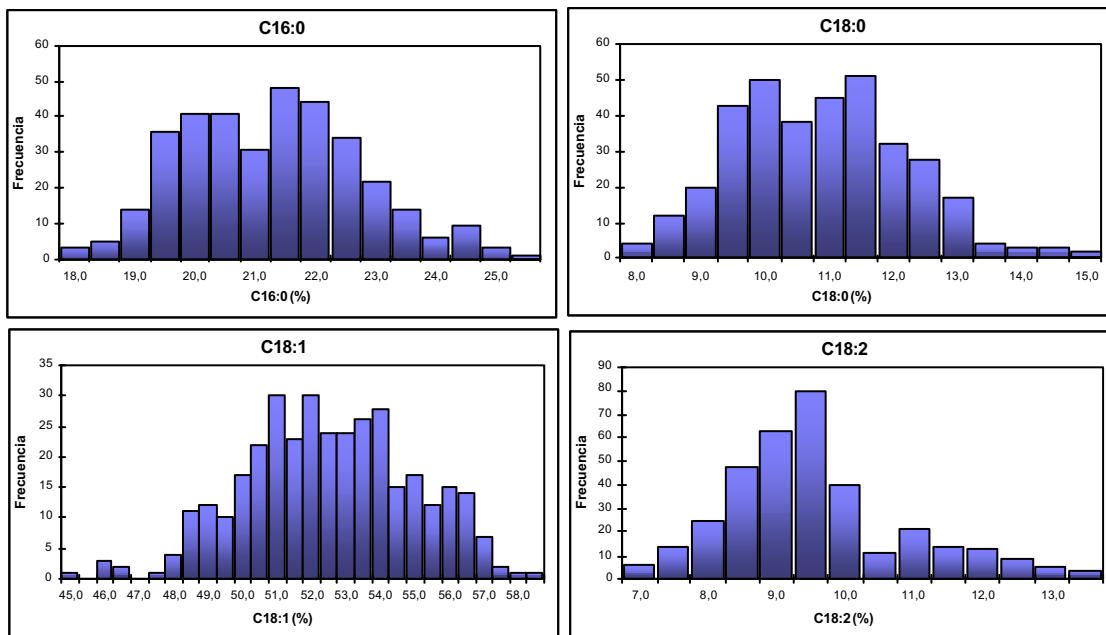
La Tabla 4.3.1 muestra la composición media, los valores máximo y mínimo y la desviación típica para cada uno de los ácidos grasos de la población de calibración. Los datos recogidos en dicha tabla junto con los datos contenidos en los diagramas de frecuencias representados en la Figura 4.3.1, indican que la selección de muestras realizada permitió cubrir prácticamente todo el intervalo de variabilidad existente en los 4 ácidos grasos para muestras de grasa de cerdo Ibérico (De Pedro, 2000).

Al mismo tiempo, se puede observar que se ha conseguido estructurar un colectivo de muestras de calibración en el cual se encuentran representados todos los valores presentes en los intervalos de composición de los datos de referencia.

Tabla 4.3.1. Composición en ácidos grasos de las muestras empleadas como colectivo de calibración (n=341).

Acido graso	Media	DT	Mínimo	Máximo
C16:0 (%)	21,08	1,46	17,90	25,30
C18:0 (%)	10,67	1,32	7,70	14,90
C18:1 (%)	52,32	2,44	45,00	58,10
C18:2 (%)	9,39	1,32	6,80	13,50

Figura 4.3.1. Diagramas de distribución de frecuencias del contenido en ácidos grasos en el colectivo de calibración global.



Una vez obtenidos los datos espectrales y químicos de las muestras del colectivo de calibración, de acuerdo con el protocolo para el desarrollo de ecuaciones de calibración NIRS descrito por Shenk y Westerhaus (1995a,b y 1996), se procedió a la evaluación de anomalos espectrales mediante el cálculo del estadístico H de cada muestra al centroide de la población.

En este caso y dado el carácter preliminar de las ecuaciones se procedió a realizar el análisis de componentes principales considerando únicamente la región espectral NIR (1.100 a 2.500 nm), dados los resultados obtenidos en el apartado 4.2.3 y considerando los espectros brutos (0,0,1,1). En el apartado 4.3.1.7 se discutirá la influencia del pretratamiento de derivadas sobre el cálculo del estadístico H para los datos espectrales de grasa fundida de cerdo Ibérico.

Tras realizar un primer análisis de componentes principales, se obtuvieron 8 espectros con una distancia H superior al valor límite ($H=3$) recomendado por Shenk y

Westerhaus (1991a). Los mayores valores de este estadístico se obtuvieron para las muestras 116 ($H=11,66$) y 98218 ($H=5,18$). Estas muestras habrían sido anómalas por la posible presencia de restos de tejido subcutáneo en el primer caso y por la presencia de burbujas en la cápsula en el segundo.

Una vez eliminados los espectros de estas muestras nuevamente se realizó un segundo análisis de componentes principales obteniéndose únicamente 3 muestras anómalas. En este caso, los valores de H de dichas muestras fueron ligeramente superiores a 3 (el máximo valor fue $H=3,68$ para la muestra 7). Tras eliminar dichas muestras se repitió de nuevo el proceso, no presentando ninguna muestra distancias H al centro de la población superiores al valor recomendado como máximo. En total, se eliminaron 11 muestras en esta etapa de detección de anomalías espectrales, las cuales fueron eliminadas del colectivo de calibración. En el apartado 4.3.1.7 se realizará una discusión más detallada acerca del origen de las anomalías espectrales de dichas muestras.

Con los datos espectrales y químicos de las muestras de dicho colectivo se procedió a la obtención de diferentes ecuaciones de calibración, para cada uno de los cuatro ácidos grasos considerados en este apartado, haciendo uso de la rutina AUTOCAL. De entre todas ellas se seleccionaron aquellas con mayor precisión y exactitud de acuerdo con los estadísticos tradicionalmente empleados en los trabajos científicos (r^2 y $ETVC$) y otros de uso menos frecuente (RPD y RER), pero de gran interés desde el punto de vista de la evaluación de la utilidad de una determinada ecuación para su aplicación real a nivel de la industria (Williams y Sobering, 1996).

La Tabla 4.3.2 muestra los estadísticos obtenidos para las ecuaciones seleccionadas como de mayor precisión y exactitud. Estas ecuaciones utilizan un tratamiento de derivadas de (1,10,5,1) para las variables C16:0, C18:1 y C18:2 y de (2,10,5,1) para el ácido graso C18:0. El número de términos RMCP seleccionados mediante validación cruzada para cada ecuación fue de 13 para C16:0, 10 para C18:0, 7 para C18:1 y 14 para C18:2.

Los resultados que se muestran en la Tabla 4.3.2 son indicadores de la alta precisión de las ecuaciones obtenidas. Así, los valores del coeficiente de determinación para los cuatro ácidos grasos estudiados fueron elevados (superiores a 0,97), tanto para

el colectivo de calibración (R^2) como para el de validación cruzada (r^2). Shenk y Westerhaus (1996) indican que las ecuaciones con valores del coeficiente de determinación superiores a 0,9 deben ser consideradas como de precisión excelente y las comprendidas entre 0,5 y 0,9 como de buena precisión.

Tabla 4.3.2. Estadísticos de ecuaciones obtenidas a partir del colectivo de calibración (n=341).

Ac. graso	Media	DT	ETL ⁽¹⁾	ETC	R ²	ETVC	R ²	RPD	RER
C16:0 (%)	21,02	1,40	0,26	0,22	0,98	0,25	0,97	5,6	33,5
C18:0 (%)	10,62	1,31	0,22	0,20	0,98	0,24	0,97	5,5	32,8
C18:1 (%)	52,29	2,33	0,25	0,21	0,99	0,24	0,99	9,7	58,3
C18:2 (%)	9,40	1,34	0,15	0,14	0,99	0,16	0,99	8,4	50,3

1: calculado a partir del análisis por duplicado de 20 muestras incluidas en la calibración

Los valores de *ETVC*, estadístico considerado como el mejor estimador que el *ETC* de la capacidad predictiva de una ecuación NIRS (Shenk y Westerhaus, 1995b), son de una magnitud similar a los errores de laboratorio (*ETL*) del método de referencia, obtenido este tras el análisis de 20 muestras por duplicado mediante el método de referencia. Ello indica no sólo la alta calidad de los datos de referencia sin asimismo la alta repetibilidad de la medida espectral y la alta correlación entre los datos espectroscópicos y los datos de referencia, responsables del bajo error del ajuste teniendo en cuenta la fórmula de partición de la varianza del error (apartado 2.2.3.1).

Al mismo tiempo, estos valores de *ETVC* son directamente comparables con los valores de desviación típica de la repetibilidad estimada en el ensayo colaborativo (apartado 4.1.3), estimador del error intralaboratorio. Los valores incluidos en la Tabla 4.1.14 de desviación típica de la repetibilidad para el ácido C16:0 (0,13), C18:0 (0,05), C18:1 (0,15) y C18:2 (0,07) fueron inferiores a los valores de *ETL* incluidos en la Tabla 4.3.2. Esto podría deberse a que en los resultados obtenidos en el ensayo colaborativo se había minimizado las fuentes de error, p. ej., empleando un inyector automático en el cromatógrafo de gases (Salas, comunicación personal). Además, se ha de recordar que los valores aportados en el ensayo colaborativo no son duplicados ciegos como los empleados en la estimación de los valores de *ETL*.

Los valores de los estadísticos *RER* y *RPD* superan los valores mínimos recomendados (10 para el *RER* y 3 para el *RPD*), confirmando de nuevo la alta exactitud

y precisión de las ecuaciones y asegurando la validez de las mismas desde el punto de vista de su aplicación a muestras futuras.

Hay que destacar también que los valores de *ETVC* reflejados en la Tabla 4.3.2, para todos los ácidos grasos, son inferiores a los obtenidos por otros autores (De Pedro et al., 1995; González, 1997) sobre muestras de grasa de cerdo Ibérico o sobre muestras de características similares (grasas animales o aceites vegetales).

Aunque estos resultados se discutirán con más detalle en el apartado 4.3.1.4, se ha de indicar a modo de resumen que, independientemente de las diferencias en instrumentación NIRS, en el tipo de cápsula y en el sistema de interacción radiación-muestra empleado, la elevada precisión de las ecuaciones obtenidas en este caso sean una lógica consecuencia, entre otros factores, de la estrategia seguida en la minimización del error espectral, aspecto no tenido en cuenta frecuentemente en el desarrollo de trabajos NIRS.

Un trabajo riguroso de optimización de la forma de presentación de muestra y de la cuantificación del error entre dos espectros de la misma muestra, y la puesta a punto de un sistema de recogida de espectros, que nos proporcionara las máximas garantías de repetibilidad de la medida espectral, todo ello unido a que los datos de referencia provienen de un laboratorio de enorme experiencia en el análisis de ácidos grasos en grasas y aceites, han contribuido, sin duda, a la minimización del error de predicción total.

Una vez obtenidas estas ecuaciones de calibración, se procedió a su evaluación mediante la predicción de muestras no empleadas en la calibración (validación externa).

Para ello se aplicaron las ecuaciones generadas sobre el colectivo de validación A, el cual contenía espectros de 20 muestras de grasa de cerdo Ibérico obtenidos justo después de finalizar la recogida de datos espectrales del colectivo de calibración. Como se observa en la Tabla 4.3.3, la composición en ácidos grasos mayoritarios de este colectivo de validación es similar, a nivel de valores medios, desviaciones típicas, máximos y mínimos, a la del colectivo utilizado para la calibración (Tabla 4.3.2).

Tabla 4.3.3. Composición en ácidos grasos del colectivo de validación A (n=20).

Ac. graso	Media	DT	Mínimo	Máximo
C16:0 (%)	20,13	1,63	17,40	23,40
C18:0 (%)	9,54	1,28	7,60	12,40
C18:1 (%)	53,56	2,70	47,40	58,10
C18:2 (%)	10,11	1,15	8,60	12,40

Los estadísticos obtenidos en el proceso de predicción de las muestras del colectivo A se indican en la Tabla 4.3.4. El valor medio del estadístico H del colectivo de validación A fue de 2,63, inferior al límite recomendado por el programa ($H=3$). Esto indica que, desde el punto de vista de las características espectrales NIRS, las ecuaciones son adecuadas para la predicción de este colectivo y que no están realizando una extrapolación del modelo de calibración (ASTM, 1994).

Por tanto, dado este valor del estadístico H y considerando los estadísticos de las ecuaciones empleadas, los valores predichos por NIRS deberían de ajustarse en gran medida a los valores obtenidos por el método de referencia.

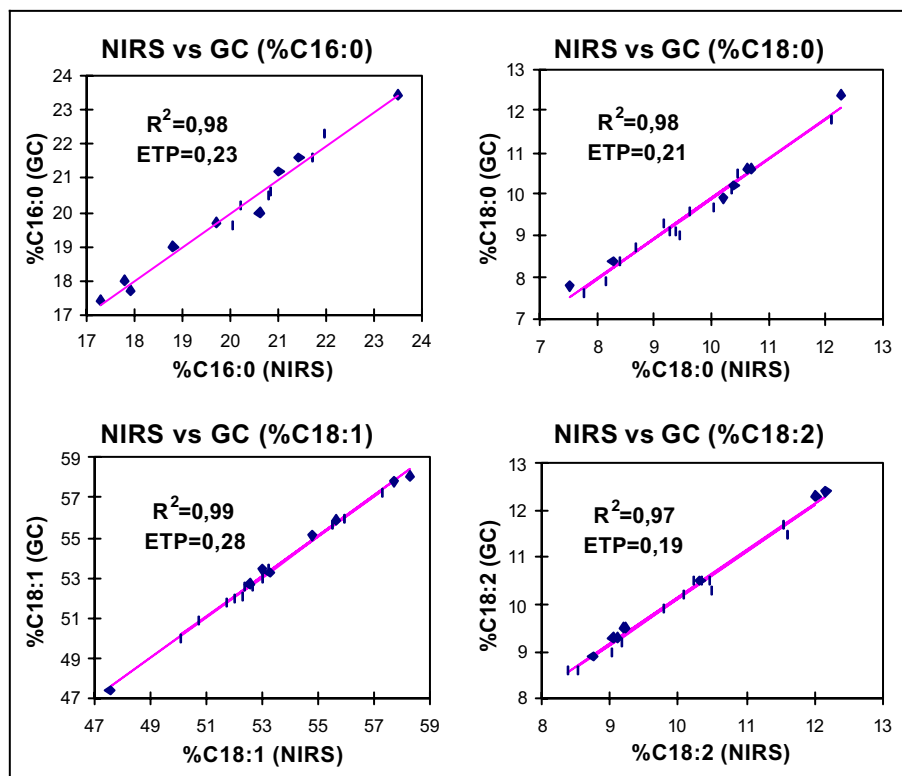
Tabla 4.3.4. Estadísticos de validación externa del colectivo A (n=20).

Ac. graso	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R ²
C16:0 (%)	0,23	-0,03	0,13	0,24	0,28	0,97	0,98
C18:0 (%)	0,21	-0,09	0,12	0,20	0,26	0,98	0,98
C18:1 (%)	0,28	0,16	0,13	0,24	0,27	1,00	0,99
C18:2 (%)	0,19	-0,03	0,08	0,19	0,18	1,00	0,97

La Figura 4.3.2 muestra las rectas de regresión de los datos predichos por NIRS frente a los datos obtenidos por cromatografía de gases, donde se observa la excelente correlación entre ambos métodos de análisis.

Un análisis de los datos de la Tabla 4.3.4, teniendo en cuenta los criterios establecidos por Shenk *et al.* (1989) y especificados en el apartado 3.3.1.4, confirma la excelente precisión de las ecuaciones obtenidas. Un hecho importante a resaltar es que el ETP o error típico de los residuales (CG *versus* NIRS) para el colectivo de validación y para cada ácido graso, es similar al ETVC (Tabla 4.3.2), lo cual pone de manifiesto lo adecuado de la utilización del ETVC como estimador del error obtenido sobre colectivos diferentes al de calibración.

Figura 4.3.2. Resultados de Validación al emplear ecuaciones NIRS generadas mediante RMCPM.



Como se observa, el *sesgo* obtenido para el ácido oleico (0,16) supera el valor límite establecido para dicho estadístico (0,13). Esto confirma, de nuevo, lo indicado por Moya *et al.* (1995) en relación a que los límites establecidos para el *sesgo* ($ETC \times 0,6$) y el $ETP(C)$ ($ETC \times 1,3$) podrían ser muy estrictos para ecuaciones de alta exactitud y precisión y por tanto bajo ETC , como es el caso de las ecuaciones utilizadas (Tabla 4.3.2) para la predicción del colectivo de validación externa.

4.3.1.2 Comparación de ecuaciones obtenidas con diferentes métodos de regresión: RMCPM vs. RLM

A partir del mismo colectivo de calibración empleado en el apartado anterior, se generaron ecuaciones de calibración NIRS para los ácidos grasos mayoritarios mediante el método de regresión lineal múltiple (RLM). Un aspecto muy importante a la hora de desarrollar este tipo de ecuaciones es fijar el número máximo de longitudes de onda por variable (ácido graso) que ha de seleccionar el algoritmo para obtener los modelos con mayor precisión. Así, se generaron ecuaciones NIRS para un máximo de 2 y 10 longitudes de onda seleccionadas por variable con la finalidad de

evaluar la influencia de este factor sobre la precisión de las ecuaciones obtenidas para un posible desarrollo de este tipo de ecuaciones en instrumentos NIRS de filtros.

En esta etapa se empleó de nuevo la rutina AUTOCAL para la obtención de ecuaciones mediante regresión RLM con diferentes tratamientos de derivadas.

La Tabla 4.3.5 muestra los estadísticos de las ecuaciones con mayor exactitud y precisión generadas mediante regresión RLM a partir de una selección máxima de 10 longitudes de onda por variable. Como puede apreciarse, los valores de R^2 obtenidos durante el proceso de calibración fueron superiores a 0,96 y los *ETC* oscilaron entre 0,16% para C18:2 y 0,27% para C18:0. El tipo de derivada empleada en las ecuaciones con menor error varió entre una 1ª derivada para el ácido C18:0, una segunda derivada para la variable C16:0 y una tercera para los ácidos C18:1 y C18:2.

Tabla 4.3.5. Estadísticos de ecuaciones de calibración generadas mediante RLM, empleando 10 longitudes de onda por variable.

Ac. graso	N	Media	DT	Derivada	ETC	R^2
C16:0 (%)	310	21,03	1,42	2,10,5,1	0,23	0,98
C18:0 (%)	319	10,64	1,33	1,4,4,1	0,27	0,96
C18:1 (%)	308	52,27	2,40	3,5,5,1	0,26	0,99
C18:2 (%)	324	9,38	1,31	3,10,10,1	0,16	0,99

Es conveniente recordar que mediante la opción incluida en el paquete estadístico ISI para la obtención de ecuaciones mediante regresión RLM, no es posible realizar una validación cruzada de las calibraciones, por ello se utilizó el *ETC* para la comparación de los resultados con los obtenidos por el método de referencia. Como puede observarse tras la comparación de estos valores con los indicados en la Tabla 4.3.2, los estadísticos de calibración obtenidos mediante regresión RLM son similares a los logrados en las ecuaciones NIRS desarrolladas empleando una regresión RMCPM.

La Tabla 4.3.6 muestra las 10 longitudes de onda seleccionadas para cada una de las variables (40 en total) en las ecuaciones NIRS generadas mediante RLM reseñadas anteriormente. Estas longitudes de onda son las que resultaron significativas para cada ácido graso tras realizar los sucesivos tests F. Como puede apreciarse en dicha tabla y en la Figura 4.3.3, la mayor parte de las longitudes de onda pertenecen a los segmentos de longitudes de onda comprendidos entre 1.100 a 1.300, 1.500 a 1.800, 2.100 a 2.300 y 2.400 a 2.500 nm., coincidiendo con las principales bandas de absorción

mostradas para la grasa de cerdo Ibérico (Figura 4.2.1) y el resto de grasas y aceites (apartado 2.3.1).

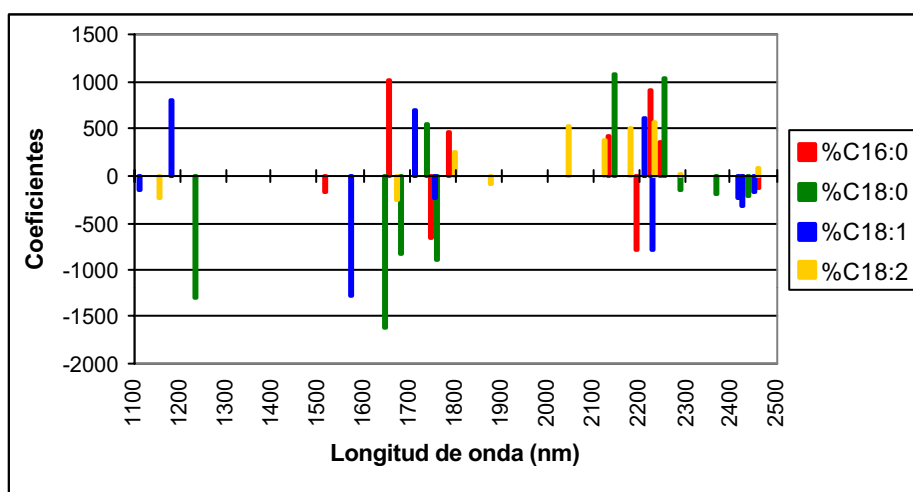
Tabla 4.3.6. Longitudes de ondas seleccionadas para ecuaciones generadas mediante RLM.

Ac. graso	Longitudes de onda seleccionadas (nm)
C16:0	1516, 1654, 1744, 1784, 2132, 2190, 2220, 2244, 2426, 2458
C18:0	1230, 1646, 1680, 1734, 1756, 2142, 2250, 2286, 2370, 2436
C18:1	1108, 1176, 1568, 1708, 1750, 2204, 2222, 2412, 2422, 2450
C18:2	1150, 1668, 1794, 1872, 2040, 2120, 2176, 2226, 2284, 2456

Cada ecuación selecciona longitudes de onda para cada variable dentro de todos los intervalos, no ciñéndose a una región en concreto. Esto indica que no se detecta ninguna región específica para la determinación de cada ácido graso sino que, de forma general, existen 3 regiones espectrales que contribuyen a dicha determinación. Al mismo tiempo, es importante destacar que dichos intervalos de longitudes de onda seleccionados mediante regresión RLM coinciden con las principales bandas de absorción de la radiación NIR por parte de la grasa de cerdo Ibérico.

En la Figura 4.3.3 se representan los valores de los coeficientes de cada longitud de onda para las ecuaciones generadas de los ácidos grasos.

Figura 4.3.3. Coeficientes de longitudes de onda seleccionadas mediante RLM en ecuaciones generadas para un máximo de 10 longitudes de onda por variable.



En ella se observa que la contribución de cada longitud de onda en un determinado ácido graso es diferente. Así, para el ácido graso C16:0, el valor del

coeficiente a la longitud de onda de 1.516 nm es mucho menor y de signo contrario (coeficiente $_{1516} = -169,57$), que la contribución de la longitud de onda a 1.654 nm (coeficiente $_{1516} = 1.018,63$). Estos valores de los coeficientes también son diferentes para los distintos ácidos grasos incluso en longitudes de ondas muy próximas. Esto sucede en el intervalo entre 2.220 y 2.250 nm, donde los cuatro ácidos grasos presentan longitudes de onda significativas si bien los valores de los coeficientes para el C16:0 (coeficiente $_{2244} = 367,62$), C18:0 (coeficiente $_{2250} = 1.031,52$), C18:1 (coeficiente $_{2222} = -771,82$) y C18:2 (coeficiente $_{2226} = 562,45$) son claramente diferentes, tanto en signo como en valor absoluto.

La Tabla 4.3.7 muestra los estadísticos de las ecuaciones con mayor precisión y exactitud generadas mediante RLM a partir de la selección máxima de 2 longitudes de onda por variable.

Tabla 4.3.7. Estadísticos de ecuaciones de calibración generadas mediante RLM, empleando 2 longitudes de onda por variable.

Ac. graso	N	Media	DT	Derivada	ETC	R ²
C16:0 (%)	327	10,07	1,44	2,10,5,1	0,42	0,92
C18:0 (%)	328	10,66	1,34	2,10,10,1	0,39	0,91
C18:1 (%)	328	52,34	2,40	3,10,5,1	0,48	0,96
C18:2 (%)	321	9,40	1,32	3,10,5,1	0,28	0,96

Como puede apreciarse, estas ecuaciones tuvieron en general menor R^2 y mayor *ETC* que aquellas obtenidas mediante regresión RMCPM, oscilando los valores de R^2 del proceso de calibración entre 0,91 para C18:0 y 0,96 para C18:1 y C18:2 mientras que para los resultados de *ETC* oscilaron entre 0,28% para C18:2 y 0,48% para C18:1. El pretratamiento de derivada seleccionada en las ecuaciones con menor error fue de segunda derivada para los ácidos C16:0 y C18:0 y tercera derivada para los ácidos C18:1 y C18:2.

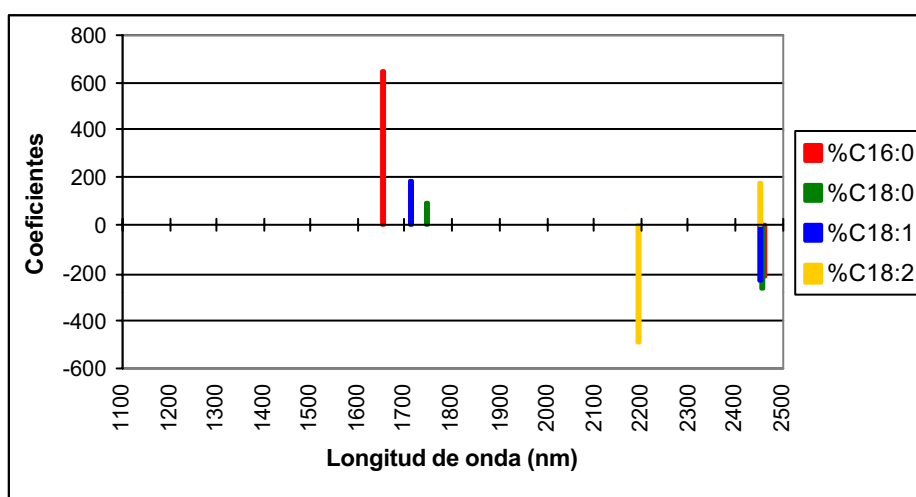
Sin embargo, las ecuaciones obtenidas pueden considerarse de una excelente precisión, dados los valores de R^2 para cada ácido graso (superiores a 0,90), los valores de *ETC* sensiblemente inferiores a los de *DT* y de magnitud inferior al doble del valor de *ETL* en todos los casos e inferiores a los errores interlaboratorios, expresados como desviación típica de la reproducibilidad (Tabla 4.1.14).

En la Tabla 4.3.8 se indican las longitudes de onda seleccionadas, tras realizar los sucesivos test F, en las 4 ecuaciones de calibración comentadas con anterioridad. Como se aprecia, las longitudes de onda seleccionadas se localizan en tres zonas muy concretas: alrededor de 1.700 nm, 2.200 nm y 2.450 nm aproximadamente (Figura 4.3.4).

Tabla 4.3.8. Longitudes de ondas seleccionadas para ecuaciones generadas mediante RLM.

Ac. graso	Longitudes de onda seleccionadas (nm)
C16:0	1.654, 2.458
C18:0	1.746, 2.456
C18:1	1.708, 2.450
C18:2	2.190, 2.450

Figura 4.3.4. Coeficientes de longitudes de onda seleccionadas mediante RLM en las ecuaciones generadas para un máximo de 2 longitudes de onda por variable.



Estas zonas de longitudes de onda coinciden en parte con las obtenidas en las ecuaciones mediante RLM con una selección máxima de 10 longitudes de onda por variable. Además, se ha de destacar que para todos los ácidos grasos se selecciona una longitud de onda en la región cercana a 2.450 nm (Figura 4.3.4), región espectral en la que presenta los picos de máxima absorción la grasa de cerdo. En este caso, de nuevo los coeficientes de las diferentes ecuaciones obtenidas son muy diferentes en valor absoluto dependiendo del ácido graso y para un mismo ácido graso de la región espectral considerada.

Una vez desarrolladas estas ecuaciones de calibración mediante regresión RLM para un máximo de longitudes de onda por ácido graso, se procedió a su validación externa empleando para ello el colectivo de validación A (Tabla 4.3.3).

La Figura 4.3.5 muestra, para cada ácido graso, las representaciones de los valores obtenidos mediante el método de referencia (cromatografía de gases) frente a los valores predichos por NIRS, a partir de las ecuaciones generadas mediante regresión RLM, empleando un máximo de 10 longitudes de onda por variable (Tabla 4.3.5). Como se puede apreciar, independientemente del ácido graso considerado, la representación de los valores se ajusta muy bien a una recta de pendiente 1, mostrando la alta precisión de las ecuaciones empleadas.

Ello se confirma al evaluar los valores de R^2 y ETP obtenidos tras la predicción del contenido en ácidos grasos mayoritarios. (Figura 4.3.5). Así, los valores de R^2 fueron superiores en todos los casos a 0,98 y los valores de ETP oscilaron entre 0,15% para C18:2 y 0,23% para C18:1. Dado que tanto los valores de R^2 como de ETP , para los 4 ácidos grasos, fueron similares a los obtenidos tras la validación con las ecuaciones generadas mediante regresión RMCPM.

Sin embargo, al aplicar ecuaciones de calibración RLM generadas a partir de 2 longitudes de onda por variable (Figura 4.3.6), los valores de ETP obtenidos superan ligeramente a los logrados anteriormente mediante ecuaciones RMCPM o RLM con 10 longitudes de onda por variable, oscilando dichos valores de ETP entre 0,21% para el ácido graso C18:2 y 0,34% para el ácido C18:0. A pesar de ello, los valores de R^2 para todos los ácidos grasos considerados fueron de nuevo muy elevados, oscilando entre 0,96 para C18:0 y 0,99 para C18:1.

Además, como se aprecia en la Figura 4.3.6, las representaciones de los valores de composición de los diferentes ácidos grasos obtenidos mediante el método de referencia (cromatografía de gases) frente a los valores predichos por NIRS mediante ecuaciones obtenidas por regresión RLM con 2 longitudes de onda por variable, se ajustan muy bien a una recta de pendiente 1. Por todo ello, se puede concluir que las ecuaciones generadas presentan una excelente precisión, si bien esta es inferior a la obtenida al emplear ecuaciones obtenidas por regresión RMCPM o por regresión RLM con un máximo de 10 longitudes de onda por variable.

Figura 4.3.5. Resultados de validación de ecuaciones NIRS generadas mediante RLM, empleando 10 longitudes de onda por variable.

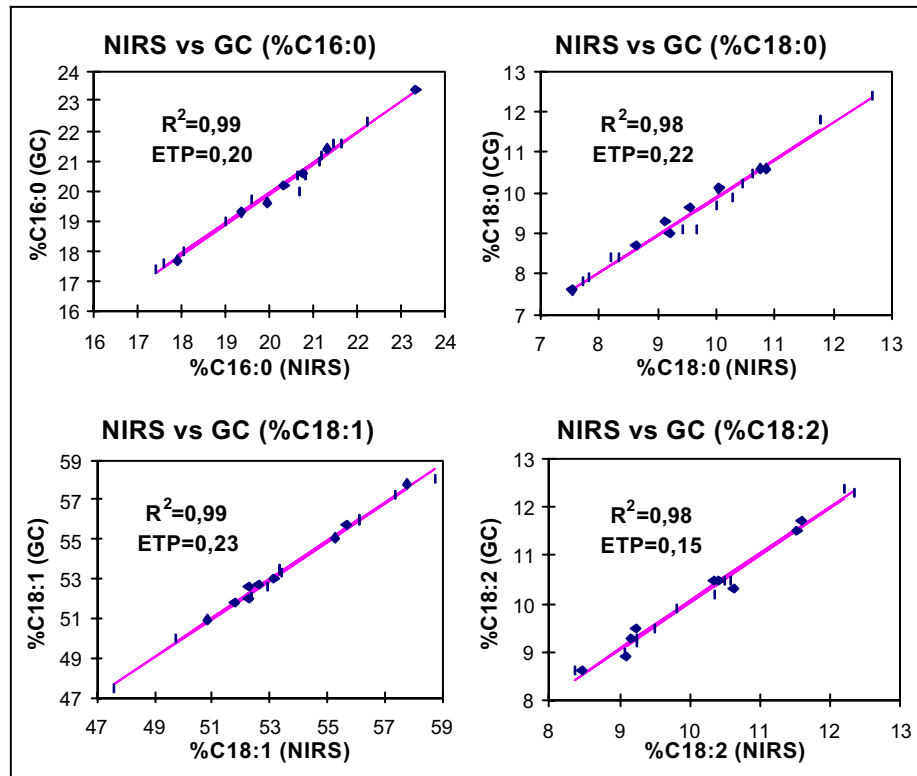
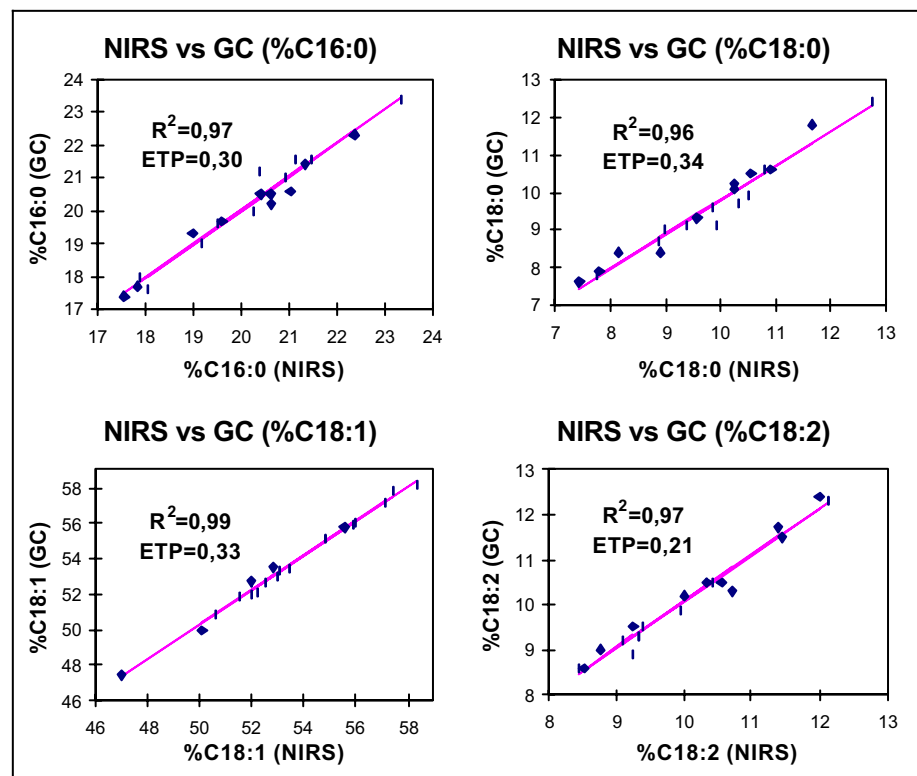


Figura 4.3.6. Resultados de validación de ecuaciones NIRS generadas mediante RLM, empleando 2 longitudes de onda por variable.



4.3.1.3 Comparación de ecuaciones obtenidas mediante diferentes tratamientos de derivadas

Como ya se indicó en el apartado 4.3.1.1, las ecuaciones preliminares desarrolladas tuvieron como principal objetivo el mostrar al sector, en concreto a la D.O. “Jamón de Huelva” las posibilidades y ventajas de aplicación de esta tecnología para el análisis de ácidos grasos en grasa de cerdo Ibérico.

Sin embargo, durante la aplicación de dichas ecuaciones se obtuvieron, para algunas muestras, elevados errores en las predicciones, aspecto este confirmado al obtener el contenido en ácidos grasos mediante el método de referencia. De acuerdo con la experiencia adquirida durante el uso de las ecuaciones las preliminares, se pensó en la posible existencia de variaciones instrumentales, operacionales o ambientales incontroladas que podrían afectar a las predicciones NIRS. Al mismo tiempo, también se detectó una cierta influencia del tratamiento de derivadas elegido durante el desarrollo de las calibraciones sobre dichas predicciones, aspecto el cual se evaluará con más detalle en este apartado.

Por todo ello, en el presente estudio se pretende estudiar la influencia del tratamiento de derivadas sobre la exactitud y precisión de diferentes ecuaciones generadas así como sobre las predicciones obtenidas para diferentes colectivos de validación externa.

Para ello se generaron ecuaciones de calibración a partir del mismo colectivo de calibración del apartado 3.3.1.1, empleando para ello el algoritmo de regresión RMCPM. Los diferentes tratamiento de derivadas, descritos de forma detallada en el apartado 3.3.1.4, consistían en 3 tratamientos diferentes de 1^a, 2^a y 3^a derivada y el correspondiente tratamiento de espectros sin derivar.

La secuencia de cálculo AUTOCAL permitió la realización de forma automática de ecuaciones de calibración modificando el tratamiento de derivadas, permaneciendo fijos el resto de variables (pretratamiento *SNV+DTR*, rango espectral entre 1.100 y 2.500 nm, límite de los estadísticos *T* y *H* de 2,5 y 10 respectivamente, un máximo de 15 términos de regresión RMCP, 4 pases de validación cruzada y número máximo de 10 pases de eliminación de muestras anómalas).

En las Figuras 4.3.7 y 4.3.8 se representan los valores de $ETVC$ y r^2 respectivamente de las 10 ecuaciones de calibración para cada ácido graso considerado, generadas mediante la secuencia automática mencionada. Como se observa, la precisión y exactitud de las ecuaciones obtenidas para todas las variables puede considerarse como excelente dados los valores de $ETVC$ y r^2 . Así, los valores de $ETVC$ son inferiores a 0,28% en todos los ácidos grasos, oscilando entre 0,21 y 0,28% para los ácidos grasos C16:0, C18:0 y C18:1 y entre 0,14 y 0,18% para el C18:2. Los valores de r^2 son superiores a 0,96 variando entre este valor y 0,99 para todos los ácidos grasos.

Figura 4.3.7. Valores de ETVC obtenidos para ecuaciones desarrolladas con diferente tratamiento de derivadas.

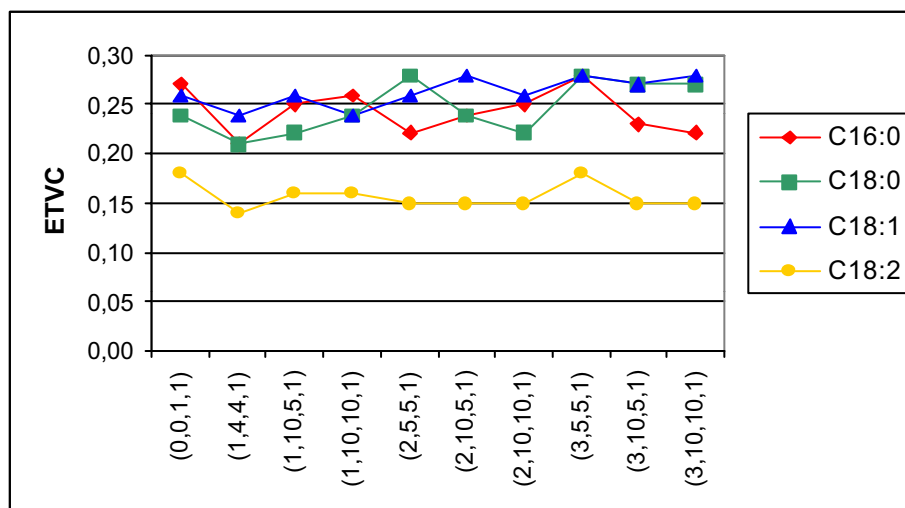
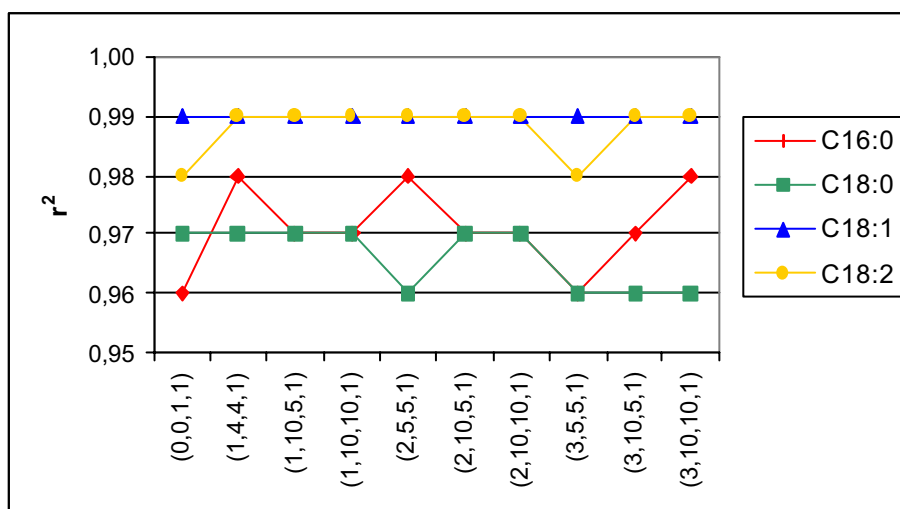


Figura 4.3.8. Valores de r^2 obtenidos para ecuaciones desarrolladas con diferente tratamiento de derivadas.



La excelente precisión y exactitud de las ecuaciones obtenidas es asimismo corroborada tras el cálculo de los estadísticos *RPD* y *RER*, los cuales se indican en la Tabla 4.3.9.

Tabla 4.3.9. Estadísticos RPD y RER de las ecuaciones desarrolladas a partir de diferentes tratamientos de derivadas.

Ecuación	Derivada	RPD				RER			
		C16:0	C18:0	C18:1	C18:2	C16:0	C18:0	C18:1	C18:2
1	0,0,1,1	5,1	5,5	8,8	7,4	30,9	32,8	52,9	44,8
2	1,4,4,1	6,6	6,2	9,7	9,6	39,8	37,1	58,3	57,4
3	1,10,5,1	5,6	6,0	8,9	8,4	33,5	35,8	53,3	50,3
4	1,10,10,1	5,3	5,5	9,6	8,4	32,1	32,6	57,8	50,3
5	2,5,5,1	6,5	4,8	9,1	8,9	39,2	28,7	54,7	53,5
6	2,10,5,1	5,9	5,5	8,4	8,8	35,4	32,8	50,6	52,7
7	2,10,10,1	5,8	6,0	8,9	8,8	34,4	35,9	53,6	52,7
8	3,5,5,1	5,1	4,8	8,6	7,3	30,6	28,8	51,6	44,1
9	3,10,5,1	6,2	4,9	8,8	8,9	36,9	29,3	52,7	53,1
10	3,10,10,1	6,5	4,8	8,4	8,7	38,6	28,9	50,5	52,1

En ella se aprecia que, para todos los ácidos grasos y tratamientos de derivadas empleados, se superan ampliamente los límites recomendados por Williams y Sobering (1996) para dichos estadísticos. Así, los valores de *RPD* para los ácidos grasos C16:0 y C18:0 oscilaron entre 5 y 6 mientras que para C18:1 y C18:2 variaron entre 7 y 9, lo cual indica un menor error de las ecuaciones obtenidas para estos últimos ácidos grasos. Esto se confirma al calcular los valores de *RER*. Así, mientras que los valores de este estadístico para los ácidos C16:0 y C18:0 oscilaron entre 30 y 40, para C18:1 y C18:2 presentó valores superiores a 50 en la mayoría de los casos.

Se ha de precisar que el número de muestras finales del colectivo de calibración, consideradas para obtener las diferentes ecuaciones de calibración, depende de la variable estudiada pero, en general, oscilaron entre un mínimo de 296 y un máximo de 330. En ningún caso, el número de muestras eliminadas como anómalas por sus valores de los estadísticos *H* y *T* supuso más del 20% del número inicial de muestras a emplear en calibración, límite recomendado por diferentes autores (Shenk y Westerhaus, 1995a) en el desarrollo de ecuaciones de calibración NIR.

Para evaluar el efecto del diferente tratamiento de derivadas sobre la precisión de las ecuaciones es necesario el comparar los diferentes estadísticos aportados por las ecuaciones. Sin embargo, dada la alta correlación existente entre los

valores predichos, es difícil poder emplear un test para comparar los estadísticos como *ETVC* o *sesgo*.

De acuerdo con Fearn (1996), para comparar el sesgo la estrategia correcta sería realizar un análisis de la varianza de los diferentes residuales y un posterior test de comparación de medias. Sin embargo, dada dicha correlación, no es posible emplear un test de comparación de varianzas del error (p.ej.: test F) para poder evaluar las diferencias entre errores del tipo del *ETVC*, *ETP* o *ETP(C)*.

Para ello, se han propuesto diferentes estrategias (Fearn, 1996; Roggo *et al.*, 2002; Næs *et al.*, 2002) si bien en este trabajo se ha empleado el criterio propuesto por Windham *et al.* (1989). Este criterio fue expuesto en un protocolo específico para el desarrollo de una ecuación de calibración NIRS y su posterior selección al considerar diferentes métodos de regresión o al modificar alguna de las variables durante su desarrollo.

Según este criterio, se han de seleccionar aquellas ecuaciones las cuales, tras realizar un proceso de validación externa, sus valores de *ETC* (ó *ETVC*) y *ETP* sean muy similares y, en cualquier caso, de debería diferir en más de un 20% uno del otro. Es decir, se fija una máxima diferencia (del orden del 20%) para considerar a ambos estadísticos similares.

Teniendo en cuenta este protocolo, se podría considerar a diferentes ecuaciones de calibración NIRS con la misma exactitud y precisión (similares en cuanto a exactitud y precisión) cuando los valores de *ETVC* no difieran entre ellos en más de un 20%. Así, se ha calculado para cada ácido graso, la diferencia máxima entre los valores de *ETVC* obtenidos de las ecuaciones con diferente tratamiento matemático a partir de la siguiente fórmula:

$$\text{Diferencia máxima } ETVC = \frac{(ETVC_{\text{máximo}} - ETCV_{\text{mínimo}})}{ETVC_{\text{mínimo}}} \times 100$$

Los valores de los *ETVC* mínimo, máximo y de la diferencia máxima calculada entre los estadísticos *ETVC* de las ecuaciones con diferentes tratamiento de derivada para cada ácido graso aparecen reflejados en la Tabla 4.3.10.

Tabla 4.3.10. Valores de ETVC máximos, mínimos y diferencias máximas, en este estadístico, para las ecuaciones de cada ácido graso.

Ac. graso	ETVC Mínimo	ETVC Máximo	Diferencia máxima ETVC (%)
C16:0 (%)	0,21	0,28	33
C18:0 (%)	0,21	0,28	33
C18:1 (%)	0,24	0,28	17
C18:2 (%)	0,14	0,18	29

Como puede apreciarse en dicha tabla, existen diferencias ligeramente superiores al valor máximo recomendado (20%) para los ácidos C16:0 (33%), C18:0 (33%) y C18:2 (29%). En el caso del ácido graso C18:1, la diferencia es del 17%, inferior al máximo recomendado. Esto muestra que el tratamiento de derivadas empleado durante la obtención de ecuaciones provocó diferencias, en general, próximas al valor límite recomendado del 20%. Se podría concluir que, desde el punto de vista de los valores de *ETVC* las diferencias entre los diferentes tratamientos de derivadas empleados, no son importantes desde el punto de vista práctico y no parecen estar asociadas a un tratamiento de derivadas concreto.

Una vez observado el efecto de las derivadas sobre el *ETVC* procedía a evaluar su influencia en un colectivo de validación externa (colectivo de validación A). La Tabla 4.3.11 muestra los diferentes estadísticos de validación externa obtenidos para los ácidos grasos y tratamientos de derivadas considerados.

De una primera observación de dicha tabla, se puede concluir que los valores de *ETP* presentan unos intervalos de variación amplios para cada uno de los ácidos grasos considerados. Estos oscilan entre valores similares a los de *ETVC*, descritos con anterioridad, hasta valores que en algunos casos llegan incluso a duplicar los valores de *ETVC* obtenidos previamente (0,22 a 0,38% para C16:0; 0,21 a 0,38% para C18:2; 0,27 a 0,48% para C18:1; 0,13 a 0,26% para C18:2).

La influencia del tratamiento de derivadas se vio reflejada asimismo en los valores de *sesgo* y, en menor medida, en los valores de *ETP(C)*, los cuales siguieron diferentes tendencias dependiendo del ácido graso considerado.

Así, para el ácido C16:0 se obtuvieron, en la mayoría de los casos, valores de *sesgos* y/o *ETP(C)* superiores a los límites recomendados por cada ecuación, siendo los valores de *sesgo* en algunos casos superiores al doble del valor límite.

Tabla 4.3.11. Estadísticos de predicción del colectivo A con ecuaciones obtenidas mediante diferente tratamiento de derivadas.

C16:0 (%)								C18:0 (%)							
Ec.	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R ²	Ec.	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R ²
1	0,30	0,01	0,15	0,31	0,33	1,00	0,96	1	0,25	-0,10	0,13	0,23	0,28	0,99	0,97
2	0,31	0,14	0,11	0,28	0,23	0,96	0,97	2	0,34	-0,29	0,10	0,18	0,22	0,99	0,98
3	0,23	-0,03	0,13	0,24	0,28	0,97	0,98	3	0,38	-0,32	0,11	0,21	0,25	0,99	0,97
4	0,22	-0,03	0,14	0,22	0,30	0,99	0,98	4	0,36	-0,29	0,13	0,21	0,27	1,00	0,97
5	0,33	-0,16	0,11	0,30	0,23	0,98	0,97	5	0,28	-0,12	0,14	0,26	0,31	0,96	0,96
6	0,38	-0,31	0,13	0,22	0,27	0,98	0,98	6	0,21	-0,09	0,12	0,20	0,26	0,98	0,98
7	0,29	-0,21	0,13	0,21	0,29	0,99	0,98	7	0,21	0,02	0,12	0,21	0,26	0,98	0,97
8	0,38	-0,01	0,13	0,39	0,28	0,94	0,95	8	0,31	-0,12	0,14	0,29	0,30	0,97	0,95
9	0,36	-0,27	0,12	0,25	0,26	0,96	0,98	9	0,22	-0,04	0,14	0,22	0,31	0,98	0,97
10	0,33	-0,21	0,12	0,25	0,26	0,97	0,98	10	0,32	-0,10	0,14	0,31	0,31	0,96	0,94

C18:1 (%)								C18:2 (%)							
Ec.	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R ²	Ec.	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R ²
1	0,44	0,36	0,14	0,26	0,30	1,01	0,99	1	0,19	-0,12	0,10	0,15	0,22	1,02	0,98
2	0,28	0,16	0,13	0,24	0,27	1,00	0,99	2	0,14	-0,02	0,07	0,15	0,14	0,98	0,98
3	0,43	0,33	0,14	0,27	0,30	0,99	0,99	3	0,19	-0,03	0,08	0,19	0,18	1,00	0,97
4	0,35	0,23	0,13	0,27	0,28	1,00	0,99	4	0,15	-0,01	0,09	0,16	0,19	1,00	0,98
5	0,37	0,29	0,15	0,24	0,31	0,99	0,99	5	0,22	0,10	0,07	0,20	0,16	1,02	0,97
6	0,48	0,44	0,15	0,18	0,33	1,00	0,99	6	0,13	-0,02	0,08	0,13	0,16	1,00	0,99
7	0,30	0,25	0,14	0,17	0,30	1,01	0,99	7	0,16	-0,01	0,08	0,16	0,17	0,99	0,98
8	0,30	0,11	0,16	0,29	0,34	0,99	0,99	8	0,26	0,07	0,09	0,25	0,20	1,04	0,95
9	0,27	0,18	0,15	0,21	0,33	1,00	0,99	9	0,14	0,00	0,08	0,14	0,17	1,00	0,98
10	0,40	0,31	0,15	0,25	0,33	1,01	0,99	10	0,14	0,03	0,08	0,14	0,17	1,02	0,99

En el caso del ácido C18:0, únicamente se apreciaron valores de *sesgo* superiores a los límites recomendados al emplear ecuaciones obtenidas con un pretratamiento de primera derivada (ecuaciones 2, 3 y 4).

Sin embargo, para el caso del ácido C18:1, los valores de *sesgo* excedieron, en mayor o menor medida, sus límites recomendados en prácticamente todas las ecuaciones mientras que en el ácido C18:2, se superaron ligeramente los límites de *sesgo* y/o *ETP(C)* con algunas ecuaciones.

Teniendo en todos los estadísticos, los resultados se pueden llegar a considerar como de adecuada precisión. Ello se confirma al evaluar los valores obtenidos de r^2 y pendiente. Los coeficientes de determinación fueron, para todas las ecuaciones y ácidos grasos considerados, superiores a 0,94 mientras que los valores de pendiente obtenidos oscilaron, para todos los casos, entre 0,94 y 1,04.

4.3.1.4 *Predicción de colectivos de validación de diferentes periodos*

Como se ha mostrado en el apartado anterior, los diferentes tratamientos de derivadas parecen dar lugar a valores de *ETP* distintos según el pretratamiento de derivadas considerado. Sin embargo, los máximos valores alcanzados por el *ETP* no aclararían los valores de error tan elevados (con residuales de incluso 2 puntos para algún ácido graso) que se habían observado en el análisis en rutina, durante la predicción del contenido en ácidos grasos en muestras de la Denominación de Origen “Jamón de Huelva”.

Un análisis detallado de las muestras cuya predicción resultó anómala durante dicho análisis en rutina indicaba que parecía existir una cierta influencia del momento de recogida del dato espectroscópico sobre el error de predicción, creando con ello la sospecha de que ciertas fuentes de error no controladas variaban entre días de análisis diferentes.

A fin de poder evaluar la robustez o estabilidad de las ecuaciones a lo largo del tiempo, se validaron las ecuaciones NIRS desarrolladas con los colectivos de validación B y C (apartado 3.3.1.1). Básicamente, estos colectivos contenían espectros de muestras de grasa de cerdo obtenidos en periodos diferentes al de recogida de los datos espectrales del colectivo de calibración.

Así, el colectivo B contenía 30 espectros, obtenidos en noviembre de 1999, correspondientes a muestras cuyos espectros fueron inicialmente recogidos en marzo de 1998 y formaron parte del colectivo de calibración. Este colectivo se diseñó para evitar incluir en el error de validación el atribuible a muestras no incluidas en la calibración. En la Tabla 4.3.12 se recoge la composición en ácidos grasos de dicho colectivo.

Tabla 4.3.12. Composición en ácidos grasos de las muestras del colectivo de validación B (n=30).

Ácido graso	Media	DT	Mínimo	Máximo
C16:0 (%)	21,08	1,98	17,90	24,60
C18:0 (%)	10,44	1,66	7,70	14,90
C18:1 (%)	52,40	3,11	46,30	57,70
C18:2 (%)	9,61	1,69	6,90	13,50

Como puede apreciarse, tanto el valor medio como la desviación típica de los ácidos grasos considerados son similares a los estadísticos del colectivo de calibración, cubriendo un intervalo de composición semejante a los de dicho colectivo (Tabla 4.3.1).

El colectivo de validación C se diseñó para contener errores no controlados del tipo instrumentales, ambientales o de preparación de muestras que puedan presentarse a lo largo de un periodo de tiempo diferente al periodo de recogida espectral del colectivo de calibración.

Dicho colectivo contenía 150 espectros de una única muestra de grasa de cerdo Ibérico (denominada *Grasa Control*). Estos 150 espectros fueron obtenidos tras el análisis, 3 veces por semana, de la muestra de grasa de referencia durante el periodo comprendido entre junio de 1999 y julio de 2000. Para estimar la composición de dicha muestra de referencia, se determinó mediante cromatografía de gases el contenido en ácidos grasos de 6 submuestras de dicha muestra de grasa a lo largo del periodo considerado. Posteriormente, estos valores fueron mediados obteniendo así la composición en ácidos grasos descrita en la Tabla 4.3.13. Como se aprecia en dicha tabla, su composición en ácidos grasos es similar a la composición media de los colectivos de calibración y de validación A y B.

Tabla 4.3.13. Composición en ácidos grasos de la muestra del colectivo de validación C (n=150).

Ácido graso	Composición
C16:0 (%)	20,91
C18:0 (%)	9,94
C18:1 (%)	52,11
C18:2 (%)	10,27

Cuando ambos colectivos (B y C) se emplearon para la validación externa de las ecuaciones iniciales desarrolladas mediante regresión RMCPM, descritas en la Tabla 4.3.2, los valores de *ETP* para ambos colectivos y ácidos grasos fueron muy superiores a los valores de *ETVC* de las ecuaciones empleadas y a los valores de *ETP* obtenidos en la predicción del colectivo de validación A partir de dichas ecuaciones preliminares (Tablas 4.3.14 y 4.3.15).

Estas diferencias llegaron a ser muy importantes en los valores predichos de C16:0 y sobre todo para el ácido graso C18:1 donde los estadísticos *ETP* de ambos colectivos son 5 veces superiores a los valores de *ETP* del colectivo A. Además, los valores medios del estadístico *H* para los colectivos B y C fueron 25,44 y 33,38 respectivamente, muy superiores al valor máximo recomendado ($H=3$). En cambio, como se ha indicado anteriormente, el valor de *H* obtenido para el colectivo de validación A fue de 2.63.

Tabla 4.3.14. Estadísticos de validación externa del colectivo B.

Ac. graso	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R ²
C16:0 (%)	0,56	-0,50	0,13	0,26	0,28	0,99	0,98
C18:0 (%)	0,28	-0,08	0,12	0,27	0,26	1,02	0,97
C18:1 (%)	1,11	1,07	0,14	0,31	0,30	1,04	0,99
C18:2 (%)	0,21	0,15	0,08	0,15	0,18	0,98	0,99

Tabla 4.3.15. Estadísticos de validación externa del colectivo C.

Ac. graso	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)
C16:0 (%)	0,55	-0,49	0,13	0,25	0,28
C18:0 (%)	0,38	0,25	0,12	0,29	0,26
C18:1 (%)	1,51	1,44	0,14	0,45	0,30
C18:2 (%)	0,44	-0,33	0,08	0,29	0,18

Esto confirmaba que pese a la excelente exactitud y precisión mostrada por las ecuaciones preliminares, algún factor o factores relacionados con el momento de análisis parece afectar de forma importante a la capacidad predictiva de dichas ecuaciones.

Al observar de forma detallada los errores presentes en la validación de los colectivos B y C, se puede apreciar que la mayor parte de estos se debe a la presencia de errores sistemáticos en todos los ácidos grasos, es decir, a la presencia de *sesgos* elevados entre los valores de referencia y los valores predichos. Estos *sesgos* superaron ampliamente los límites máximos para dichos estadísticos en las ecuaciones obtenidas, fundamentalmente en las ecuaciones de C18:1. Tan solo en el colectivo B, la ecuación de predicción del contenido en C18:0 presentó un sesgo (0,08%) ligeramente inferior al límite (0,12%).

Sin embargo, los valores de $ETP(C)$ fueron en todos los casos similares e incluso en algunos ácidos grasos inferiores a los límites máximos aceptados en función de la precisión de las ecuaciones. Por tanto, se puede apreciar que la mayor parte del error existente en la validación de los colectivos B y C se debió a errores sistemáticos (evaluados por el *sesgo*), siendo prácticamente inexistente la presencia de errores aleatorios (evaluados por el estadístico $ETP(C)$). Ello se confirma con los valores de pendiente y R^2 existente entre los valores de referencia y predichos del colectivo B (para el colectivo C no tiene sentido hablar de dichos estadísticos al existir un único valor de referencia). Dichos estadísticos muestran una excelente precisión, con valores próximos a 1 en ambos casos.

La presencia de errores en la predicción llama la atención dada la alta exactitud y precisión de las ecuaciones desarrolladas. Este tipo de errores se puede explicar tras evaluar las diferencias entre los diferentes colectivos de validación y más en concreto su periodo de análisis espectral. Así, los espectros incluidos en el colectivo A se obtuvieron justo a continuación de la toma de espectros del colectivo de calibración, por lo que las condiciones instrumentales y ambientales fueron similares en el análisis de ambos grupos de muestras.

Sin embargo, los colectivos de validación B y C se obtuvieron 1 y 2 años después que el colectivo de calibración, por lo que las condiciones instrumentales y de análisis fueron diferentes. Esto se corrobora al evaluar los valores medios del estadístico H obtenidos para cada colectivo de validación. Así, mientras que en el colectivo de validación al valor de H fue inferior al límite recomendado, para los colectivos B y C este estadístico superó ampliamente dicho límite, mostrando con ello que los espectros de ambos colectivos eran muy diferentes (anómalos) a los del colectivo de calibración.

Conocida la nula influencia de la conservación y manipulación de las muestras sobre el valor de este estadístico (apartado 4.2.3) y teniendo en cuenta que las muestras del colectivo B y C son similares en cuanto a origen y composición al colectivo de calibración (para el colectivo B son incluso las mismas muestras) los factores que debieron provocar los elevados valores de ETP podrían estar ligados a fuentes de variación instrumentales y/o ambientales. En el apartado siguiente se tratará de profundizar en la evaluación de dichas variaciones instrumentales y/o ambientales.

4.3.1.5 Control y diagnósticos de rutina de la instrumentación NIRS utilizada

Dada la importancia que, presumiblemente, podrían llegar a tener los factores instrumentales y/o ambientales sobre la precisión de las ecuaciones, se decidió incluir como parte del trabajo de desarrollo y evaluación de calibraciones un estudio específico de evaluación de los distintos diagnósticos del instrumento a lo largo de todo el periodo de recogida espectral considerado en esta Tesis Doctoral.

Afortunadamente, los diagnósticos diarios y semanales que el Servicio Centralizado NIR-Banco de Muestras Valoradas (posteriormente Unidad NIR/MIR del Servicio Central de Apoyo a la Investigación) de la Universidad de Córdoba realizaba al instrumento Foss NIRSystems SYI desde su adquisición en 1991 permitían un estudio detallado de dichos diagnósticos (apartado 3.3.3).

El periodo considerado para el estudio de los resultados de los diferentes test de diagnósticos diarios y semanales ha sido de 2 años y 10 meses; concretamente desde el mes de octubre de 1997 al mes de julio de 2000 (ambos inclusive). A lo largo de este periodo se ha realizado la toma de espectros de muestras de grasa de cerdo Ibérico de diferente composición y procedencia las cuales, o bien formaron parte del colectivo de calibración o bien fueron empleadas como colectivos de validación externa (A, B y C).

A continuación, se indican los sucesos más relevantes acontecidos durante dicho periodo, los cuales permitirán explicar posteriormente algunos problemas detectados durante la recogida de espectros realizada para esta Tesis:

- **6/02/98:** Se detecta que la lámpara del instrumento está fundida y se reemplaza por una nueva. Es necesario ajustar la respuesta del instrumento (al tener un valor superior a 65.000) y corregir los valores de K y ϕ en el rango NIR, al detectarse un valor elevado en el error (superior a 0,56 nm) de estima de la exactitud de la longitud de onda de dicho rango de medida.
- **19/05/98:** se cambia la lámpara del instrumento a pesar de no estar fundida debido a la presencia de valores muy altos (máximo >34 $\mu\log$) en el ruido espectral o

repetibilidad de la señal analítica NIR. Es necesario una corrección de los valores de K y ϕ en el rango NIR.

- **24/03/99:** Valores muy altos (máximo $>32 \mu\log$) en la repetibilidad de la señal analítica NIR. Se procede al cambio de la lámpara del instrumento por una nueva aun a pesar de no dar signos de que estuviera fundida. Este cambio no disminuye el nivel de ruido por lo que se decide continuar con la lámpara antigua y, aunque se desconoce él por qué, los valores de ruido recuperan sus niveles habituales.
- **19/05/99:** Se detecta la entrada de luz al instrumento cuando realiza la medida de la cerámica (referencia interna). Este hecho afectó de forma muy notable a la repetibilidad de la señal analítica NIR (máximo $>30 \mu\log$). Al corregir la posición de la cerámica, desaparecen los altos niveles de ruido del instrumento.
- **5/04/00:** Se detectan altos valores de error (superior a $0,30 \text{ nm}$) en la exactitud de las longitudes de onda en el rango NIR. Por ello, se corrigen los valores de K y ϕ para dicho rango.

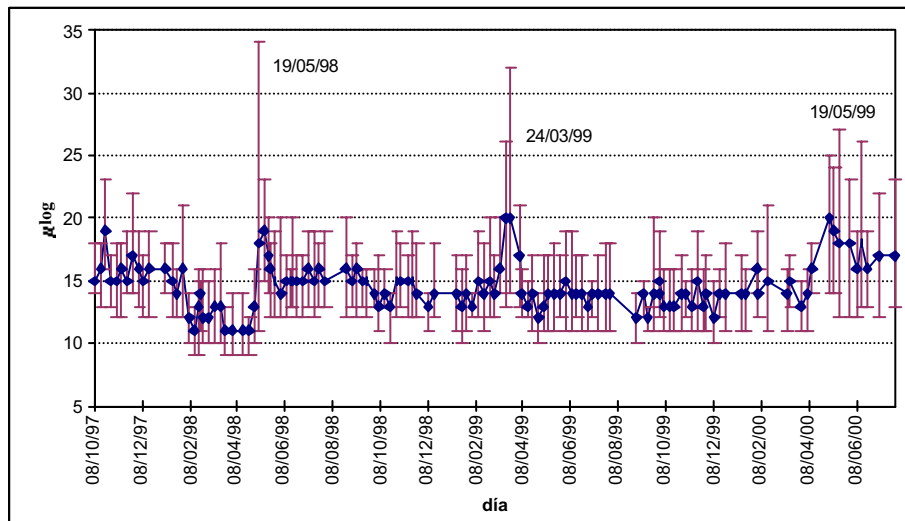
Los resultados de realizar cada uno de los test de diagnósticos (repetibilidad de la señal analítica NIR, exactitud de la longitud de onda y respuesta de los detectores del instrumento), se describen a continuación.

En la Figura 4.3.9 se muestran los valores de $RMS(C)$ máximos, mínimos y medios (realizados sobre 10 espectros) de cada uno de los test de repetibilidad de la señal NIR realizados a lo largo del periodo estudiado.

Únicamente se han evaluado los resultados debidos a la repetibilidad de la señal analítica en el rango NIR, al ser este el único rango considerado en el desarrollo de las ecuaciones de calibración para determinar el contenido en ácidos grasos en grasa fundida de cerdo Ibérico.

Como se observa en dicha figura, los valores medios de $RMS(C)$ no superan en ningún caso el valor límite recomendado de $20 \mu\log$ ($1/R$). Sin embargo, los valores máximos de $RMS(C)$ si superan ampliamente este límite en varias ocasiones del periodo considerado. Esto ocurrió fundamentalmente en tres fechas determinadas: mayo de 1.998, marzo de 1.999 y mayo de 2.000.

Figura 4.3.9. Diagnostico de ruido. Valores medios, máximo y mínimos de $RMS(C)$, para el rango NIR, en el periodo de tiempo estudiado.



En el primer caso (mayo de 1.998) la presencia de valores de $RMS(C)$ elevados pudo deberse a defectos de fabricación de la lámpara ya que no cabe esperar que se debiera a un deterioro de la lámpara por el uso ya que sólo llevaba 3 meses en funcionamiento y en los 10 años de uso del instrumento, se ha observado una vida media de la lámpara de 12 meses (Baena, comunicación personal).

En el segundo caso (abril de 1.999) se sospechó de nuevo de defectos de lámpara al obtenerse valores muy elevados de ruido. Sin embargo, tras sustituir la lampara existente por una nueva, no se observó disminución en los valores de $RMS(C)$, descartándose el factor lámpara como fuente de error.

Se detectó la entrada de radiación externa a los detectores durante el desarrollo del test de repetibilidad. Esta radiación procedía de la lámpara halógena existente en la sala la cual incidía sobre los detectores debido a una incorrecta posición de la cerámica de referencia. Esta entrada de luz “parásita”, como la denomina Williams (1987), es una fuente de error importante. Una vez suprimida dicha entrada de luz, los niveles de ruido volvieron a sus niveles habituales.

En el tercer caso (mayo de 2.000), los altos valores de $RMS(C)$ coincidieron en el tiempo con errores elevados en la determinación de la exactitud de las longitudes de onda, lo cual provocó, presumiblemente, altos niveles de ruido. Tras mejorar la exactitud, corrigiendo los valores K y ϕ para ambos rangos disminuyeron los valores

medios de $RMS(C)$ si bien aun se detectaron valores máximos superiores al límite recomendado.

La evaluación de la exactitud de las longitudes de onda se ha realizado a partir de los valores de error existentes entre las longitudes de onda nominales del poliestireno y las longitudes de onda obtenidas por el instrumento en cada uno de los controles semanales.

Se ha calculado este error para el mismo periodo y rango de medida espectral considerado en el test de repetibilidad. A pesar de que los límites recomendados para el error son el no superar el doble del valor del error sugerido al realizar el test, una recomendación análoga es que dicho valor no sea superior a 0,30 nm (apartado 3.3.3).

La Figura 4.3.10 muestra los valores de error en la estima de la exactitud en la medida de las longitudes de onda para el rango NIR. Se puede apreciar que las principales oscilaciones del error, debidas a correcciones de los valores de K y ϕ en ambos rangos, fueron motivadas por cambios de lámpara (6/02/98, 19/05/98 y 24/03/99).

La última disminución del error apreciada se corresponde con una corrección de los valores de K y ϕ realizada el 5/04/00. Dicha corrección fue consecuencia de una tendencia al alza de los valores del error hasta valores cercanos a 0,30, situación lógica y normal en un instrumento NIR tras su elevado uso y sobre el que no se había realizado ninguna corrección de valores de K y ϕ para el rango NIR en los últimos 12 meses.

En la Figura 4.3.11 se muestra la evolución de la respuesta del instrumento a lo largo del periodo considerado para el rango NIR. Así, el valor de la respuesta osciló alrededor del valor considerado como óptimo (55.000) con 3 grandes oscilaciones provocadas por los cambios de lámpara (6/02/98, 19/05/98 y 24/03/99) y con ligeras correcciones de la respuesta a lo largo de todo el periodo debido a pequeñas pérdidas en la señal de los detectores.

Figura 4.3.10. Exactitud de longitud de onda para el rango NIR en el periodo de tiempo estudiado.

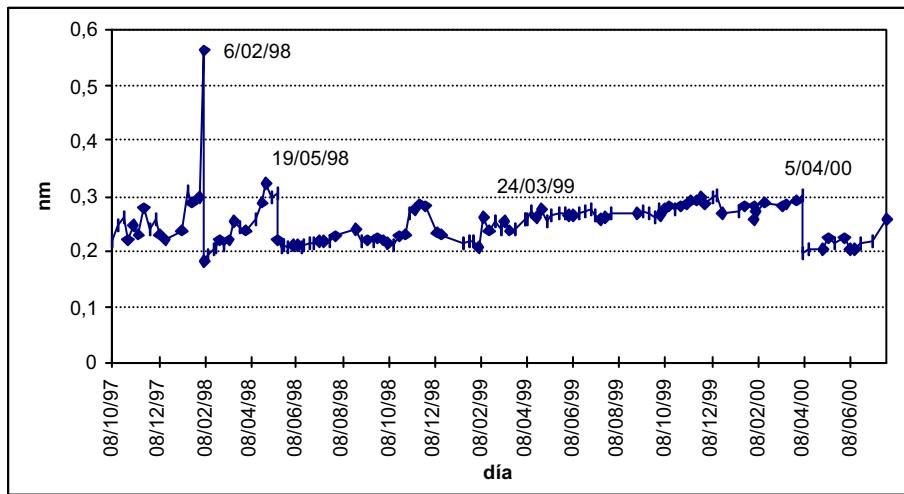
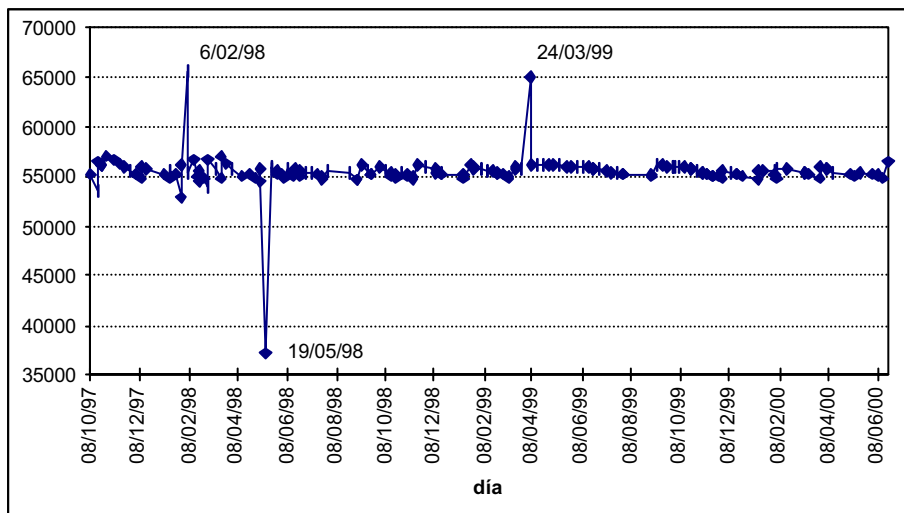


Figura 4.3.11. Respuesta de la lámpara para el rango NIR en el periodo de tiempo estudiado.



Junto a estos diagnósticos de rutina del instrumento, se realizó el análisis de control diario de la muestra de referencia (*Check Cell*). De acuerdo con lo descrito en el apartado 3.3.3.4, en la Tabla 4.3.16 se recogen los valores estimados de \bar{y}_{mr} y σ_{mr} para los constituyentes de la muestra de referencia, a partir de las predicciones realizadas en su análisis por primera vez en el instrumento Foss NIRSystems 6500 SY I (junio de 1.994) y al comienzo de la toma de espectros de grasa de cerdo Ibérico (octubre de 1.997).

Tabla 4.3.16. Valores de \bar{y}_{mr} y σ_{mr} para los constituyentes de la muestra de referencia.

Constituyente	\bar{y}_{mr}	σ_{mr}
Proteína bruta (% MS)	57,33	0,10
Grasa bruta (% MS)	2,28	0,02
Fibra bruta (% MS)	4,85	0,04
Materia seca (%)	90,57	0,02

A partir de estos valores y para un nivel de significación del 99%, los límites de confianza para cada constituyente vendrán definidos por la fórmula ($\bar{y}_{mr} \pm 3 \times \sigma_{mr}$). Así, los límites de confianza serán 57,33% \pm 0,30 para la proteína bruta, 2,28% \pm 0,06 para la grasa bruta, 4,85% \pm 0,12 para la fibra bruta y 90,57% \pm 0,06 para la materia seca.

En las Figuras 4.3.12 a 4.3.15 se han representado en un gráfico de control los valores de proteína bruta, grasa bruta, fibra bruta y materia seca predichos para la muestra de referencia dentro del periodo de desarrollo de modelos cuantitativos y cualitativos para grasa de cerdo Ibérico (de octubre de 1.997 a julio de 2.000). Como fue descrito en el apartado 3.3.3.4, en dicho periodo, la muestra de referencia fue analizada como una muestra de composición desconocida. Junto a dicho valores también se han representado los valores de \bar{y}_{mr} (en verde) y los límites de confianza (en rojo) para cada constituyente.

De acuerdo con la norma ASTM E 1655 (1994), cabe esperar que los valores predichos de la muestra de control se encuentren dentro de los límites de confianza definidos previamente para cada constituyente en un 99% de los casos. Esto ocurre para los valores de grasa bruta y materia seca de la muestra de referencia, como se muestra en las Figuras 4.3.13 y 4.3.15.

Sin embargo, en el caso de proteína (Fig. 4.3.12) y fibra bruta (Fig. 4.3.14) existe un cierto número de valores que se ubican fuera de sus límites de confianza, apreciándose una clara deriva de la media decreciente para la proteína y creciente para la fibra bruta. En ambos casos, estos valores se obtuvieron en el periodo comprendido entre octubre de 1.998 y marzo de 1.999. A partir de este periodo, los valores vuelven a ubicarse dentro de los límites de confianza y desaparece la deriva de la media para los valores de proteína bruta, permaneciendo en el caso de la fibra bruta.

Figura 4.3.12. Predicciones de Proteína Bruta (% en MS) en muestra de referencia (Check Cell) durante el periodo considerado.

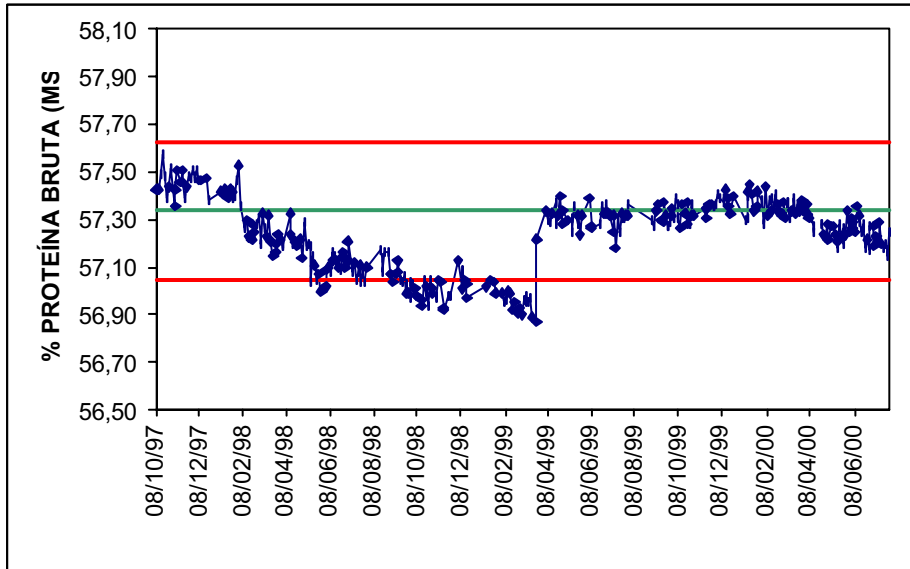


Figura 4.3.13. Predicciones de Grasa Bruta (% en MS) en muestra de referencia (Check Cell) durante el periodo considerado.

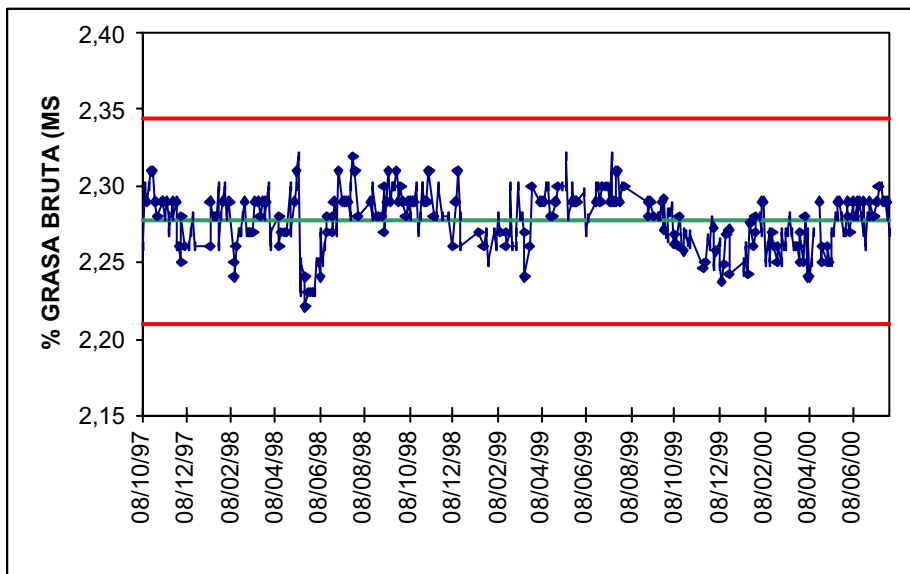


Figura 4.3.14. Predicciones de Fibra Bruta (% en MS) en muestra de referencia (Check Cell) durante el periodo considerado.

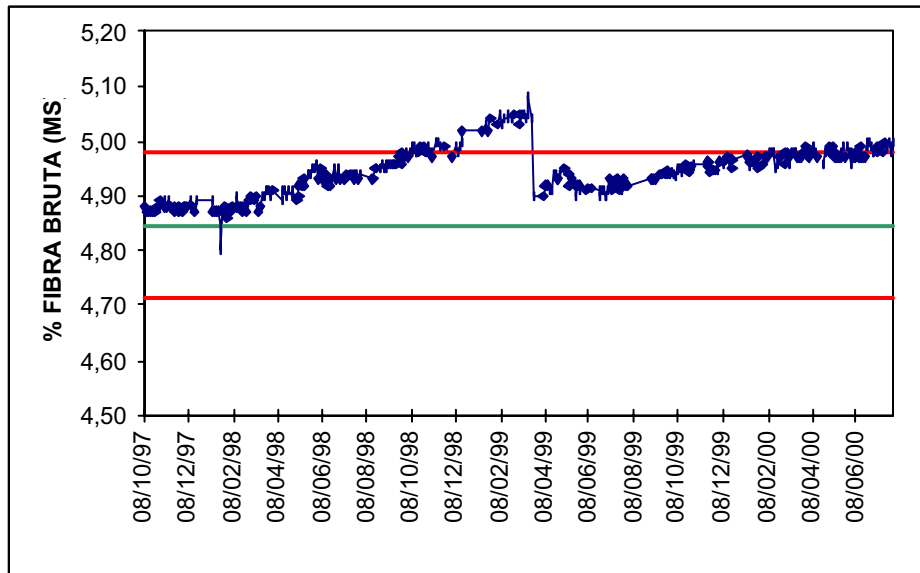
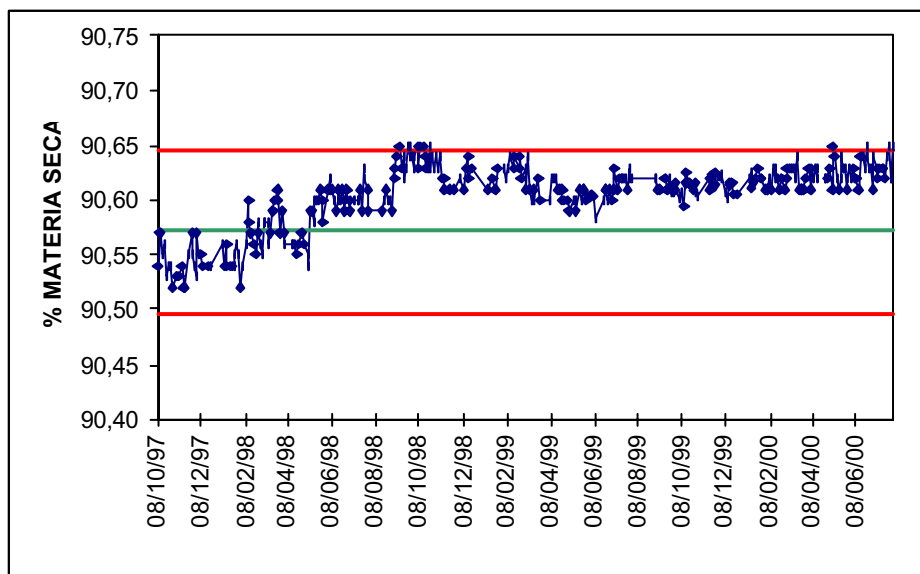


Figura 4.3.15. Predicciones de Materia Seca (%) en muestra de referencia (Check Cell) durante el periodo considerado.



Este cambio brusco de las predicciones para la proteína y la fibra bruta tuvo lugar el día 25/3/99, coincidiendo en fecha con la instalación de una lámpara nueva

debido a la existencia de errores elevados en la repetibilidad del instrumento y la reposición de la antigua tras comprobar que la lámpara no era la fuente de dichos errores. Como consecuencia de esta reposición, se produjeron unas correcciones en los valores de K y ϕ , fundamentalmente para el rango del Visible, los cuales fueron, previsiblemente los que motivaron los cambios bruscos en esta fecha (25/3/99) de las predicciones para la proteína y fibra bruta.

Al mismo tiempo, se ha de indicar que, como previene la norma ASTM (1994), los límites de confianza calculados pueden llegar a ser demasiado restrictivos. Ello se debe a que estos límites de control se basan en la repetibilidad de las predicciones en base a los espectros y no presentan una marcada influencia de los errores de calibración del modelo y del método de referencia. A esto hay que añadir que dicha muestra de referencia presenta un cierto grado de alteración en el tiempo que presumiblemente origine cambios en su composición.

Todos estos factores provocan que los límites de confianza recomendados por el propio fabricante y suministrador de las muestras de referencia sean menos restrictivos. Así, se recomienda unos límites de confianza iguales a los calculados según la norma ASTM (1994) para el caso de la proteína bruta ($\pm 0,30$ calculado vs. $\pm 0,30$ recomendado por fabricante), ligeramente superiores para el caso de grasa bruta y materia seca ($\pm 0,06$ calculado vs. $\pm 0,10$ recomendado por fabricante) y más del doble en el caso de fibra bruta ($\pm 0,12$ calculado vs. $\pm 0,30$ recomendado por fabricante).

Tras considerar estos nuevos límites de confianza, todos las predicciones para los parámetros grasa bruta, fibra bruta y materia seca permanecerían dentro de ellos. Para el caso de la proteína bruta, los límites de confianza deberían de ser algo superiores (aproximadamente $\pm 0,40$) para lograr que todos los valores se ajustaran a ellos.

Por tanto se demuestra cómo factores fácilmente detectables tales como una rotura de lámpara, la entrada de luz exterior al instrumento, y otros factores no conocidos pueden afectar a la repetibilidad de la señal espectroscópica, a la exactitud de la medida de la longitud de onda, a la respuesta de los detectores del instrumento y a los valores predichos de la muestra control (*Check Cell*).

Este conocimiento nos será de utilidad posteriormente para la interpretación de los resultados de los valores de predicción obtenidos en los colectivos de validación B y C.

4.3.1.6 Aplicación de ecuaciones obtenidas con diferentes tratamientos de derivadas a colectivos de validación de diferentes periodos

Una vez confirmado el diferente comportamiento de las ecuaciones de calibración desarrolladas en 4.3.1.1 sobre los colectivos de validación A, B y C mostrando como dicho comportamiento puede ser atribuido a factores de variación instrumental y/o ambientales y tras evaluar la instrumentación NIRS utilizada a lo largo del periodo de estudio, resta el proponer una metodología de trabajo que permita mantener en el tiempo la alta exactitud y precisión de las calibraciones.

Teniendo en cuenta las variaciones en los valores del *ETP* obtenidos al predecir el colectivo A con ecuaciones que difieren en el pretratamiento de derivadas utilizado (Tabla 4.3.11), se pretendió evaluar si dichos pretratamientos afectaban de igual forma producían también un efecto diferente en la exactitud y precisión de las predicciones de los colectivos B y C.

Para ello, se procedió a validar los colectivos B y C con las ecuaciones desarrolladas en el apartado 4.3.1.3 tras la aplicación de la secuencia automática AUTOCAL. En las Figuras 4.3.16 a 4.3.19 se aprecian los diferentes valores de *ETP* obtenidos a partir de las ecuaciones desarrolladas con diferentes tratamientos de derivadas para cada uno de los ácidos grasos y para los colectivos B y C. Se recoge asimismo los valores de *ETP* anteriormente obtenidos para el colectivo A (Tabla 4.3.11) para facilitar la discusión.

Tras una primera observación de dichas gráficas, se puede apreciar que los valores de *ETP* tras emplear ecuaciones con diferente pretratamiento de derivadas para todos los ácidos grasos mayoritarios con los colectivos B y C son, en la mayoría de las ocasiones, sensiblemente superiores a los valores obtenidos con el colectivo de validación A.

Así, los valores de *ETP* obtenidos con el colectivo A no superan el valor de 0,50% para los ácidos C16:0, C18:0 y C18:1 y de 0,25% para C18:2, mientras que los *ETP* obtenidos con los colectivos B y C sí superan, en la mayoría de los casos, el valor de 1,00% para los ácidos C16:0, C18:0 y C18:1 y de 0,50% para C18:2.

Figura 4.3.16. Valores de ETP de diferentes colectivos con ecuaciones desarrolladas para C16:0 con diferente tratamiento de derivadas.

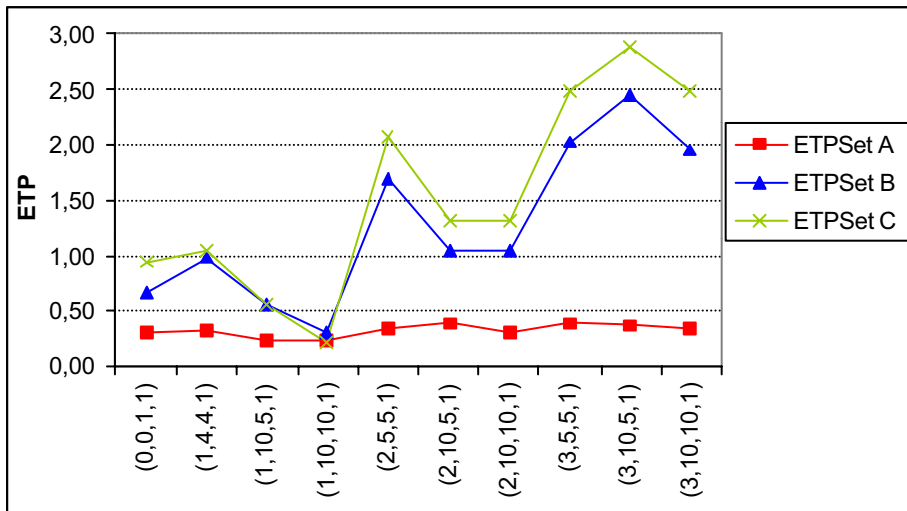


Figura 4.3.17. Valores de ETP de diferentes colectivos con ecuaciones desarrolladas para C18:0 con diferente tratamiento de derivadas.

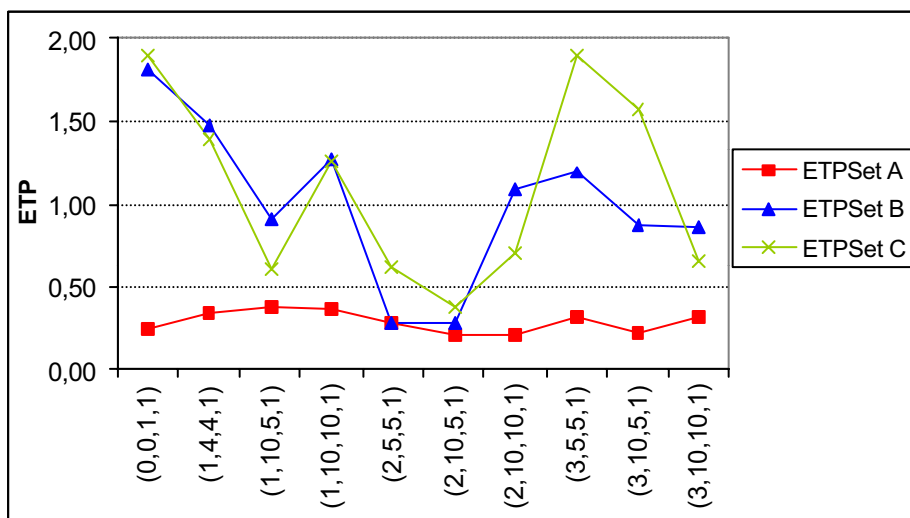


Figura 4.3.18. Valores de ETP de diferentes colectivos con ecuaciones desarrolladas para C18:1 con diferente tratamiento de derivadas.

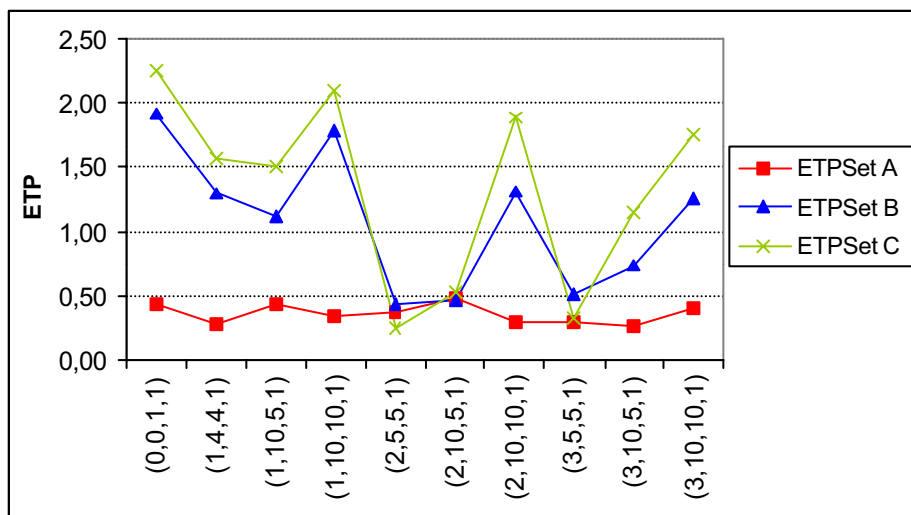
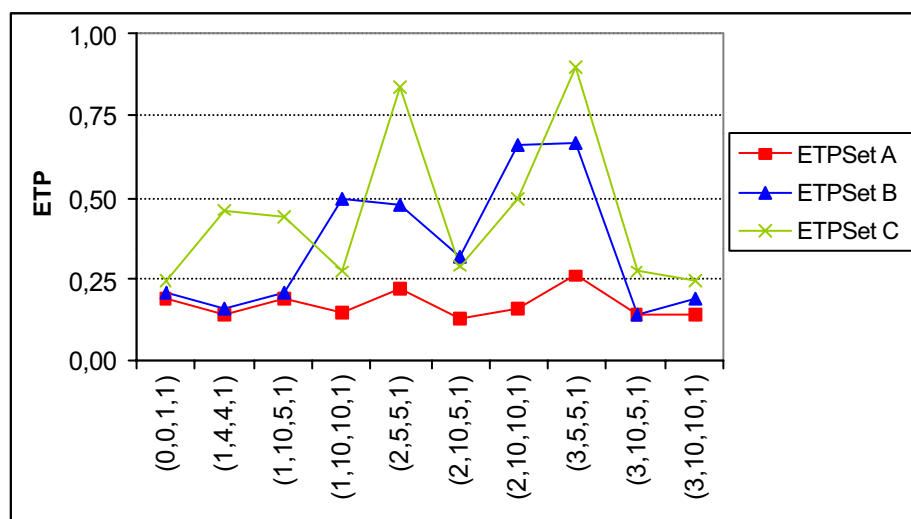


Figura 4.3.19. Valores de ETP de diferentes colectivos con ecuaciones desarrolladas para C18:2 con diferente tratamiento de derivadas.



Además, dentro de los valores de *ETP* obtenidos para cada ácido graso en los colectivos B y C, se observan claras diferencias entre dichos valores dependiendo de la ecuación (tratamiento de derivadas) empleada, si bien cada ácido graso presentó un comportamiento diferente. Estas diferencias fueron del orden de 2 puntos entre los valores máximos y mínimos de *ETP* para los ácidos C16:0, C18:0 y C18:1 y de 0,75 para C18:2. Es decir, el *ETP* de las validaciones de los colectivos B y C varió marcadamente con el tipo de tratamiento de derivadas empleado en el desarrollo de las ecuaciones, oscilando entre valores de *ETP* similares a los obtenidos con el colectivo de validación A hasta valores 5 veces superiores.

Este hecho puede deberse a la diferente robustez presentada por las distintas ecuaciones desarrolladas. Así, mientras unas ecuaciones presentaban errores muy superiores a los admisibles en función de sus valores de *ETVC* provocados presumiblemente por diferencias instrumentales y/o ambientales, otras lograban minimizar al máximo dichas diferencias obteniendo valores de *ETP* similares a sus *ETVC*. Es decir, la elección de un tratamiento de derivadas provocaría que la ecuación pueda ser considerada más robusta al validar con precisión colectivos espectrales tomados en varios periodos bajo condiciones instrumentales y/o ambientales diferentes.

Esto se confirma al apreciarse una tendencia similar, dentro cada ácido graso, de la evolución del *ETP* frente a las ecuaciones obtenidas con diferente tratamiento de derivadas. Así, para el caso del C16:0, se pueden apreciar en la Figura 4.3.16 tres máximos locales en la representación del valor de *ETP* de los colectivos B y C frente al tipo de tratamiento de derivadas empleado: en los tratamientos (1,4,4,1), (2,5,5,1) y (3,10,5,1). Algo similar ocurre para el ácido C18:0 (Figura 4.3.17) y C18:1 (Figura 4.3.18) si bien los máximos locales se aprecian en tratamientos de derivadas diferentes. Ello significaría que los diferentes factores instrumentales ya comentados y/o diversos factores ambientales podrían estar afectando de forma similar a los colectivos de validación B y C. Además, dichos factores podrían verse corregidos o minimizados por la elección adecuada de un determinado tratamiento de derivadas durante el desarrollo de la ecuación.

Por este motivo, y dada la mayor precisión y robustez observada en algunos tratamientos matemáticos, se seleccionaron aquellas ecuaciones desarrolladas con los tratamientos matemáticos de derivadas que poseían mejores estadísticos de validación. En la Tabla 4.3.17 se describen los estadísticos de calibración de las ecuaciones seleccionadas.

Tabla 4.3.17. Estadísticos de ecuaciones obtenidas a partir del colectivo de calibración (n=341).

Ac. graso	Derivada	Media	DT	ETC	R ²	ETVC	r ²	RPD	RER
C16:0 (%)	1,10,10,1	21,00	1,39	0,23	0,97	0,26	0,97	5,3	32,1
C18:0 (%)	2,10,5,1	10,62	1,31	0,20	0,98	0,24	0,97	5,5	32,8
C18:1 (%)	2,5,5,1	52,24	2,37	0,24	0,99	0,26	0,99	9,1	54,7
C18:2 (%)	3,10,10,1	9,39	1,30	0,13	0,99	0,15	0,99	8,7	52,1

Como puede apreciarse, dichos estadísticos son similares a los descritos en las ecuaciones preliminares (Tabla 4.3.2), aumentando ligeramente los valores de *ETVC* para las ecuaciones de C16:0 (de 0,25 a 0,26%) y C18:1 (de 0,24 a 0,26%) y permaneciendo constante e incluso disminuyendo para las ecuaciones de C18:0 (0,24% en ambos casos) y C18:2 (de 0,16 a 0,15%). Además, en las ecuaciones descritas en la Tabla 3.3.17 destacan los valores de *ETC* similares a *ETVC*, los bajos valores de *ETVC* y los altos valores de r^2 , de *RPD* y *RER*.

En este punto, han de ser comparados los estadísticos de calibración de este trabajo con aquellos obtenidos por otros autores para similares tipos de muestras como grasa animal fundida o aceites vegetales, ya descritos en el apartado 2.3.2.

Así, De Pedro *et al.* (1992) obtuvo valores de *ETVC* superiores a los expuestos anteriormente (0,46% para C16:0; 0,49% para C18:0; 0,56% para C18:1 y 0,33% para C18:2). La mejora en los valores de *ETVC* de las ecuaciones generadas en esta Tesis Doctoral respecto a los obtenidos por De Pedro *et al.* (1992) puede deberse al uso de un mayor número de muestras así como a una optimización del tratamiento matemático a emplear durante el desarrollo de las ecuaciones de calibración.

González (1997) obtuvo valores de *ETC* para grasa de cerdo Ibérico de 0,83% para C16:0, 0,74% para C18:0, 1,11% para C18:1 y 0,51% para C18:2. Parte de las diferencias observadas entre los resultados indicados por González (1997) y los recogidos en la Tabla 4.3.17 pueden ser debidos al uso de diferente tipo de instrumentos (Foss NIRSystems 6500 *vs.* Bran+Luebbe IA500) así como al diferente modo de análisis (doble transmisión *vs.* reflectancia mediante una cápsula abierta). Sin embargo, sí parece apreciarse que las grandes diferencias observadas en los valores de *ETVC*, en particular para alguno de los ácidos grasos (p.ej.: C18:1) que es cuatro veces superior (1,11%) al obtenido en el presente trabajo (0,26%) pueda interpretarse, probablemente, por las diferencias existentes en *ETL* para el análisis de referencia.

Chen y Chen (1995), mediante un instrumento NIRSystems 6500 y empleando la modalidad de análisis de transmisión, obtuvieron valores de R^2 y *ETP* de 0,86 y 1,21% para C18:1 y 0,83 y 1,36% para C18:2, a partir de muestras de aceite de soja mezclado con aceites de cacahuete, sesamo y girasol. Los autores concluyeron que los valores de *RPD* fueron 7,2 y 6,8 respectivamente, siendo dichos valores superiores a

los valores recomendados por Williams y Sobering (1993) para este estadístico con el fin de considerar a una ecuación como de elevada precisión.

Los valores del estadístico *RPD* obtenidos para las 2 ecuaciones de dichos ácidos grasos generadas en este trabajo fueron 9,8 y 7,3 respectivamente. Estos mejores valores de *RPD* pueden ser interpretados por las mayores desviaciones típicas para C18:1 y C18:2 en colectivo de calibración. Ha de destacarse que en su trabajo, Chen y Chen (1995) emplearon regresión lineal múltiple (RLM) y únicamente 3 longitudes de onda (1.720, 1.750 y 2.150 nm) para el desarrollo de las ecuaciones de calibración NIRS. Sin embargo, como fue descrito en el apartado 4.3.1.2. es posible obtener modelos mediante RLM con un máximo de 2 ó 10 longitudes de onda por variable para predecir el contenido de ácidos grasos en grasa de cerdo Ibérico con una precisión similar a la de los modelos obtenidos mediante regresión RMCPM.

Pérez Vich *et al.* (1998) mediante un instrumento Foss NIRSystems 6500 y analizando las muestras de aceite sobre discos de fibra óptica mediante la modalidad de reflectancia, desarrollaron ecuaciones de calibración para determinar el contenido en ácidos grasos en aceite de girasol. Los valores de r^2 obtenidos para C16:0, C18:0, C18:1 y C18:2 oscilaron también entre 0,97 y 0,99. Estos altos valores de r^2 se pueden interpretar debido a que los autores emplearon, como colectivo de calibración, un amplio número de muestras de aceite extraído de semillas de girasol provenientes de un programa de mejora genética, con lo cual lograron cubrir un gran rango de variabilidad de contenido en ácidos grasos.

Sin embargo, y teniendo en cuenta los valores medios de composición en ácidos grasos del colectivo de calibración, los valores de *ETVC* son superiores a los indicados en la Tabla 4.3.17. Ello podría ser debido a la menor repetibilidad espectral que presentan aquellos datos espectroscópicos obtenidos mediante discos de fibra de vidrio frente a los obtenidos mediante la modalidad de doble transmisión.

Ha *et al.* (1998) generaron ecuaciones en un instrumento de filtros (Bran+Luebbe IA450) y en un instrumento monocromador (Bran+Luebbe IA500) para la predicción de ácidos grasos en aceite de sésamo. Los valores de R^2 y *ETP* logrados para las ecuaciones con mayor precisión fueron de 0,82 y 0,33% para C16:0, 0,78 y 0,14% para C18:0, 0,93 y 0,52% para C18:1 y 0,92 y 1,04% para C18:2. En función de

los datos recopilados en la bibliografía (Bockisch, 1998) acerca de los valores medios de C16:0 (8,9%), C18:0 (4,8%), C18:1 (39,3%) and C18:2 (41,3%) presentes en aceite de sésamo, se puede observar que los errores obtenidos por Ha *et al.* (1998) son superiores a los indicados en la Tabla 4.3.17.

Pérez Marín *et al.* (2002), empleando la metodología para el análisis de grasas mediante la modalidad de doble transmisión propuesta en el presente trabajo de Tesis Doctoral, obtuvieron unos valores para los estadísticos de *ETVC* (0,28% para C16:0, 0,27% para C18:0, 0,20% para C18:1 y 0,16% para C18:2) y R^2 (0,94 para C18:2, 0,97 para C16:0 y 0,99 para C18:0 y C18:1) del mismo orden a los reflejados en la Tabla 4.3.17.

Garrido *et al.* (2002) desarrollaron ecuaciones para la determinación de varios parámetros de interés en grasas destinadas a alimentación animal, utilizando la metodología propuesta en este trabajo. Los valores de *ETVC* de las ecuaciones generadas para los ácidos grasos mayoritarios (0,76% para C16:0, 0,74% para C18:0, 0,97% para C18:1 y 0,65% para C18:2) fueron ligeramente superiores a los reflejados en la Tabla 4.3.17. Los valores de r^2 descritos por Garrido *et al.* (2001) fueron en todos los casos inferiores a los obtenidos en la Tabla 4.3.17.

A modo de resumen, indicar que las ecuaciones de calibración NIRS para la determinación de ácidos grasos en grasa desarrolladas mostraron una alta exactitud y precisión frente a aquellas desarrolladas por otros autores lo cual podría ser debido a varios motivos tales como el uso de un alto número de muestras que abarcan toda la variabilidad en cuanto a composición de ácidos grasos del producto se refiere, el uso de espectros con alta repetibilidad debido al uso de un valor límite de RMS para duplicados espectrales previo mediado, el uso de el modo de análisis de doble transmisión y también, probablemente, por la optimización del tratamiento matemático a emplear durante el desarrollo de las calibraciones.

Para describir con más detalle estas ecuaciones, en las Figuras 4.3.20 a 4.3.23 se exponen los valores de los coeficientes de las ecuaciones descritas en la Tabla 4.3.17. Dichos coeficientes son los valores que relacionan, mediante una combinación lineal, las variables X originales (valores de absorbancias a cada longitud de onda) con cada variable Y (ácido graso) a predecir.

Una primera observación de dichas figuras muestra que prácticamente todas las longitudes de onda, para todos las ecuaciones, presentan valores de coeficientes distintos de cero, por lo que contribuyen con signo positivo o negativo en la predicción del ácido graso a considerar.

Figura 4.3.20. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación seleccionada para determinar C16:0.

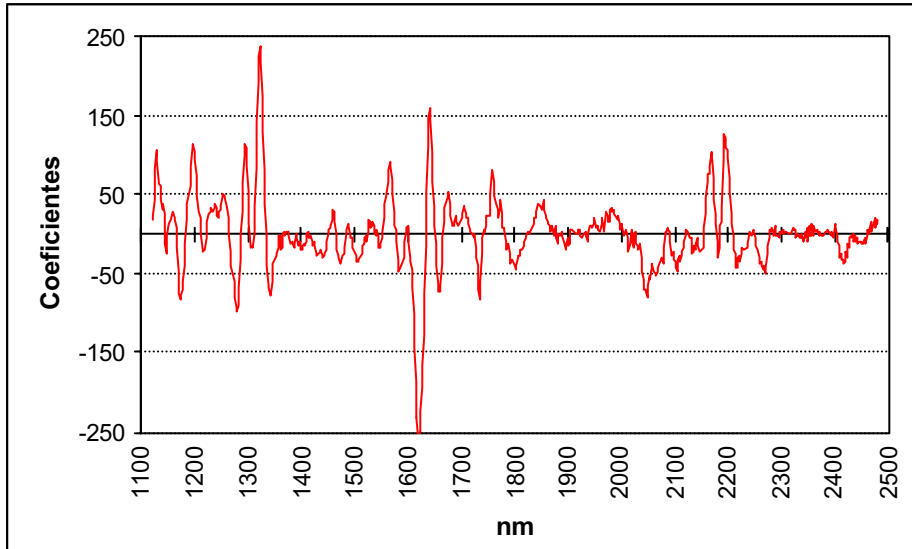


Figura 4.3.21. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación seleccionada para determinar C18:0.

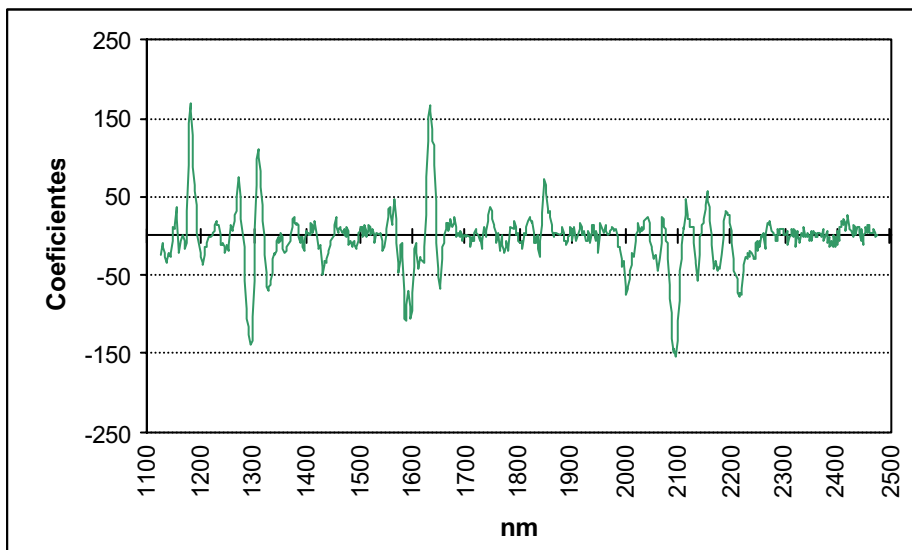


Figura 4.3.22. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación seleccionada para determinar C18:1.

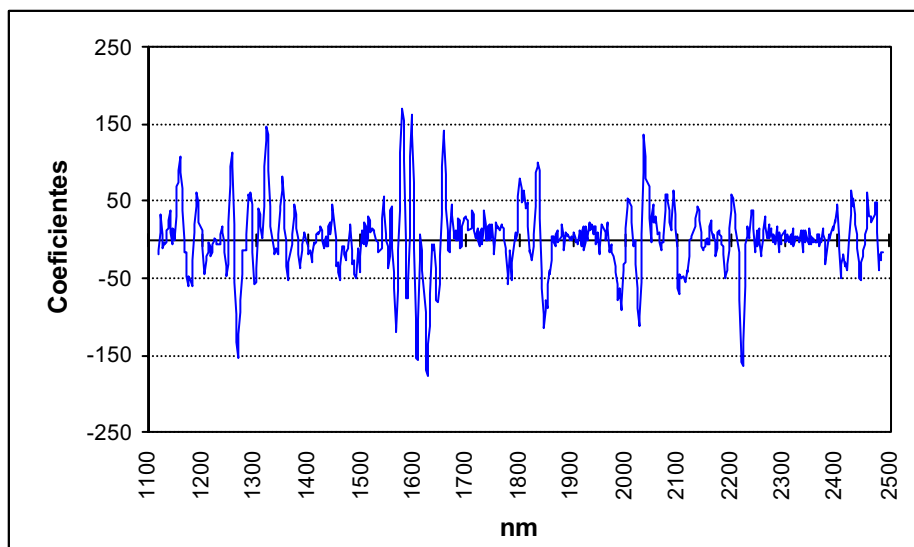
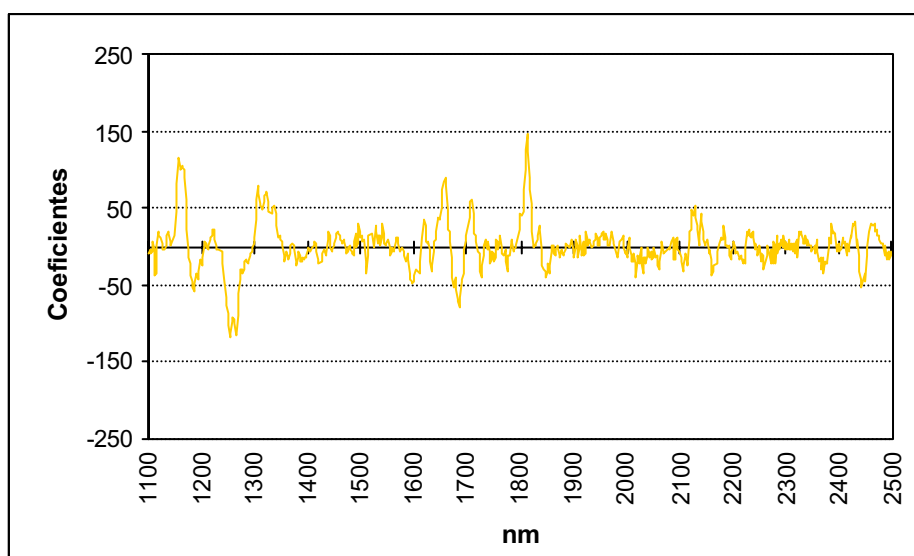


Figura 4.3.23. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación seleccionada para determinar C18:2.



Sin embargo, se puede apreciar que en todas las ecuaciones, los valores máximos y mínimos de dichos coeficientes, se encuentran localizados en determinadas regiones del espectro. En concreto, estas zonas que tendrían una mayor aportación en los valores de ácidos grasos a predecir son de 1.200 a 1.300 nm, entre 1.600 y 1.700 nm y alrededor de 2.200 nm, las cuales coinciden con picos de absorción del espectro de grasa. Sin embargo, cabe resaltar que la región en la que los espectros de grasa presentan una máxima absorción (entre 2.300 y 2.400 nm) no presentan valores de

coeficientes elevados en ninguna de las ecuaciones expuestas por lo que parece confirmar los resultados de Sato *et al.* (1991) quienes indican como región espectral idónea la incluida entre 1.100 y 2.200 nm.

Los estadísticos de la predicción de los colectivos de validación A, B y C mediante el uso de las ecuaciones seleccionadas anteriormente aparecen reflejados en las Tablas 4.3.18 a 4.3.20.

Tabla 4.3.18. Estadísticos de predicción del colectivo A con mejores ecuaciones.

Ac graso	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R ²
C16:0 (%)	0,22	-0,03	0,14	0,22	0,30	0,99	0,98
C18:0 (%)	0,21	-0,09	0,12	0,20	0,26	0,98	0,98
C18:1 (%)	0,37	0,29	0,15	0,24	0,31	0,99	0,99
C18:2 (%)	0,14	0,03	0,08	0,14	0,17	1,02	0,99

Tabla 4.3.19. Estadísticos de predicción del colectivo B con mejores ecuaciones.

Ac graso	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R ²
C16:0 (%)	0,29	-0,13	0,14	0,26	0,30	1,01	0,98
C18:0 (%)	0,28	-0,08	0,12	0,27	0,26	1,02	0,97
C18:1 (%)	0,43	-0,33	0,15	0,27	0,31	0,99	0,99
C18:2 (%)	0,19	0,12	0,08	0,16	0,17	0,98	0,99

Tabla 4.3.20. Estadísticos de predicción del colectivo C con mejores ecuaciones.

Ac graso	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)
C16:0 (%)	0,21	-0,01	0,14	0,21	0,30
C18:0 (%)	0,38	0,25	0,12	0,29	0,26
C18:1 (%)	0,25	-0,05	0,15	0,24	0,31
C18:2 (%)	0,24	-0,02	0,08	0,24	0,17

Las predicciones del colectivo A para los diferentes ácidos grasos (Tabla 4.3.18) siguen teniendo una excelente precisión, siendo similares e incluso inferiores (a excepción del C18:1) los valores de *ETP* a los obtenidos mediante las ecuaciones iniciales (Tabla 4.3.4). Únicamente se sobrepasan los límites recomendados para el *sesgo* en el caso del ácido C18:1 (0,29% frente a un límite de 0,15%), hecho que ya ocurrió en la predicción de este colectivo mediante las ecuaciones iniciales. A pesar de

ello, los valores de $ETP(C)$ para todos los ácidos grasos son inferiores a los límites máximos recomendados por el programa. Los valores de pendiente y R^2 para todos los ácidos grasos son próximos a la unidad.

Las predicciones de los colectivos B (Tabla 4.3.19) y C (Tabla 4.3.20) para los diferentes ácidos grasos presentaron, como era de esperar, unas claras mejoras en precisión. Así, los valores de ETP disminuyeron sensiblemente frente a los expuestos en las Tablas 4.3.14 y 4.3.15, fundamentalmente para los ácidos C16:0 y C18:0.

A pesar de ello y de forma puntual, se obtuvieron algunos valores de *sesgo* y $ETP(C)$ para los ácidos grasos estudiados, superiores a los límites recomendados. En la mayoría de las ocasiones, dichos valores fueron similares a los límites (*sesgo* para C18:2 y $ETP(C)$ para C18:0 en el caso del colectivo B y $ETP(C)$ para C18:0 y C18:2 en el colectivo C). Sin embargo, en el caso de los ácidos C18:1 para el colectivo B y del ácido C18:0 para el colectivo C, los valores de *sesgo* obtenidos duplicaron los valores límites establecidos para dichos estadísticos. Este hecho de nuevo podría explicarse según lo indicado por Moya *et al.* (1995) en relación a que los límites establecidos tanto para el *sesgo* como para el $ETP(C)$ pueden llegar a ser muy estrictos para ecuaciones de alta precisión, y por tanto bajo ETC , a la hora de predecir colectivos mediante validación externa.

Una vez conocido el efecto del pretratamiento de derivadas sobre la capacidad predictiva de las ecuaciones, cuando estas se aplican a colectivos espectrales análogos obtenidos en periodos diferentes al colectivo de calibración, y seleccionados aquellos tratamientos de derivación que mostraron menores valores de ETP para todos los ácidos grasos y colectivos estudiados (Tabla 3.3.17), se evaluó la repetibilidad de los valores de ácidos grasos obtenidos mediante dichas ecuaciones.

Para ello se utilizó la metodología propuesta por la norma ASTM (1994) basada en la evaluación de los errores obtenidos al predecir una muestra a lo largo de un periodo de tiempo. En este caso, se utilizó el colectivo C de espectros de la muestra de *Grasa Control*. En las Figuras 4.3.24 a 4.3.27 se muestran los gráficos de control, herramienta muy empleada en control de calidad, de las predicciones obtenidas para los diferentes ácidos grasos del colectivo de validación C, a lo largo de 13 meses.

En dichas figuras se representa, mediante una línea verde horizontal, el valor de composición para cada ácido graso en la muestra de grasa de cerdo de referencia (Tabla 4.3.13). A partir de este valor y teniendo en cuenta la norma ASTM (1994) se pueden definir unos límites de confianza (LC) para el valor estimado, mediante las ecuaciones NIRS, de composición de la muestra de grasa de control. Así, los valores predichos mediante una ecuación NIRS para una submuestra de la muestra de grasa de referencia han de pertenecer al intervalo definido por los respectivos límites de confianza.

Según dicha norma, los límites de confianza para un valor estimado mediante un modelo multivariante (en este caso las ecuaciones NIRS generadas mediante regresión RMCPM) vendrían dado por la expresión:

$$LC = \bar{y}_m \pm (t \times ETC \times \sqrt{1+h})$$

donde \bar{y}_m es el valor medio estimado para cada ácido graso en la muestra de grasa de referencia (Tabla 4.3.13), t es el valor de la t de Student para los grados de libertad del modelo y nivel de probabilidad considerado, ETC es el error estándar de calibración y h el estadístico *leverage*, análogo a la distancia de Mahalanobis. Teniendo en cuenta que el valor que puede presentar el estadístico h para muestras anómalas no suele ser superior a 0,1 y que es más adecuado el uso del error estándar de validación cruzada ($ETVC$) como estimador del error de predicción de una ecuación, la expresión anterior puede quedar simplificada a:

$$LC = \bar{y}_m \pm (t \times ETVC)$$

Dependiendo del valor considerado para el estadístico t de Student, es decir, del nivel de probabilidad de rechazar muestras ajenas a la población, se pueden definir 2 tipos de límites de confianza. Así, teniendo en cuenta el elevado número de grados de libertad y considerando el nivel de significación del 5%, el valor de t de Student se puede aproximar a 2. En dicho caso, los límites de confianza se denominan límites de aviso (ASTM, 1994). Para las ecuaciones seleccionadas (Tabla 4.3.17), los límites de aviso estimados serían 20,91%±0,52 para el ácido C16:0, 9,94%±0,48 para C18:0, 52,11%±0,52 para C18:0 y 10,27%±0,30 para C18:2. Estos límites aparecen reflejados en las Figuras 4.3.24 a 4.3.27 como líneas amarillas horizontales.

Si el nivel de probabilidad considerado para rechazar valores es del 1% y el modelo posee un elevado número de grados de libertad, el valor de t de Student se puede aproximar a 3. En dicho caso, los límites de confianza se denominan límites de actuación (ASTM, 1994). Para las ecuaciones descritas en la Tabla 4.3.17, los límites de actuación estimados serían $20,91\% \pm 0,78$ para el ácido C16:0, $9,94\% \pm 0,72$ para C18:0, $52,11\% \pm 0,78$ para C18:0 y $10,27\% \pm 0,45$ para C18:2. Dichos límites se muestran en las Figuras 4.3.24 a 4.3.27 como líneas rojas horizontales.

Figura 4.3.24. Gráfico de control de valores predichos para C16:0 en colectivo C.

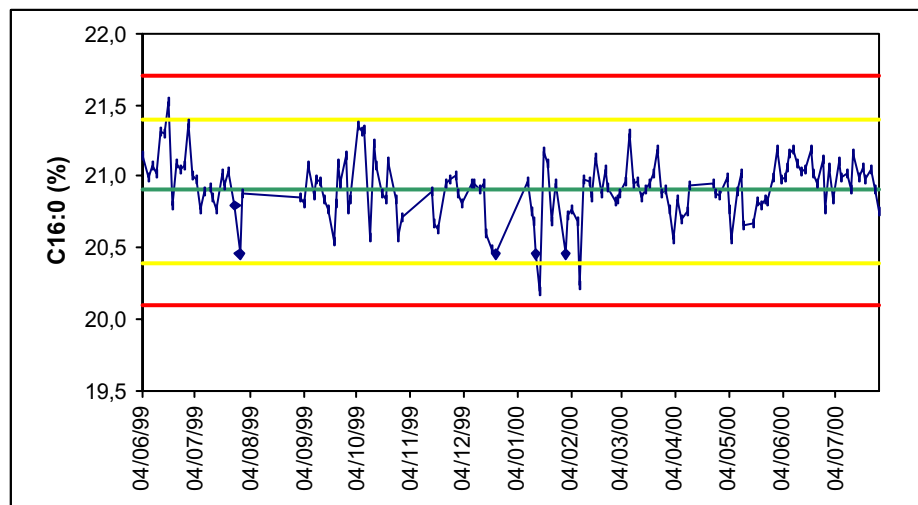


Figura 4.3.25. Gráfico de control de valores predichos para C18:0 en colectivo C.

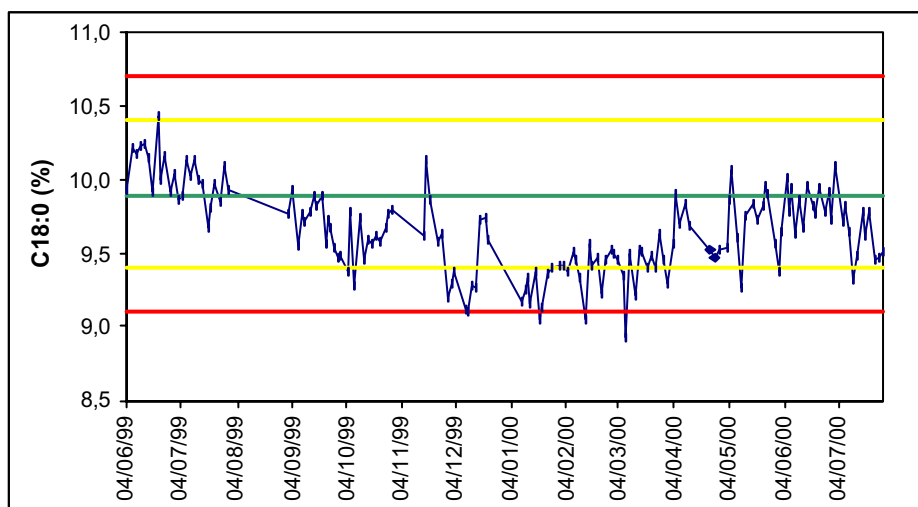


Figura 4.3.26. Gráfico de control de valores predichos para C18:1 en colectivo C.

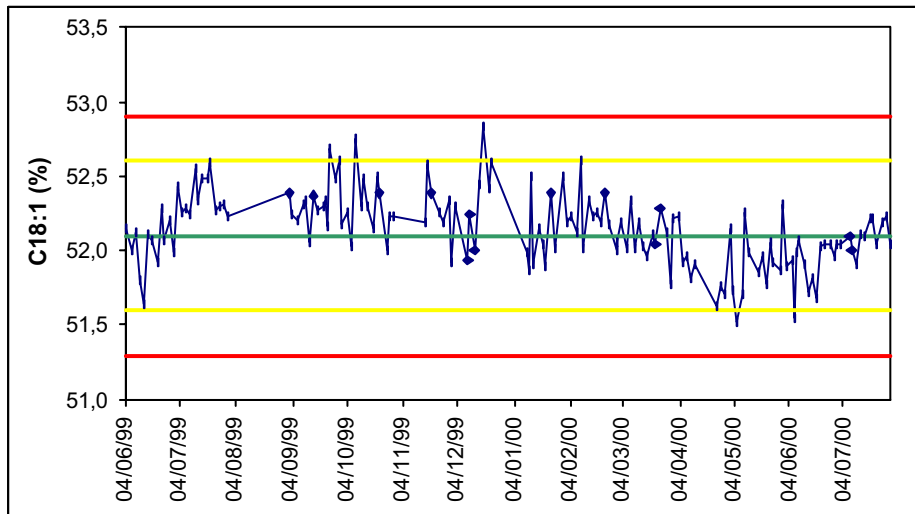
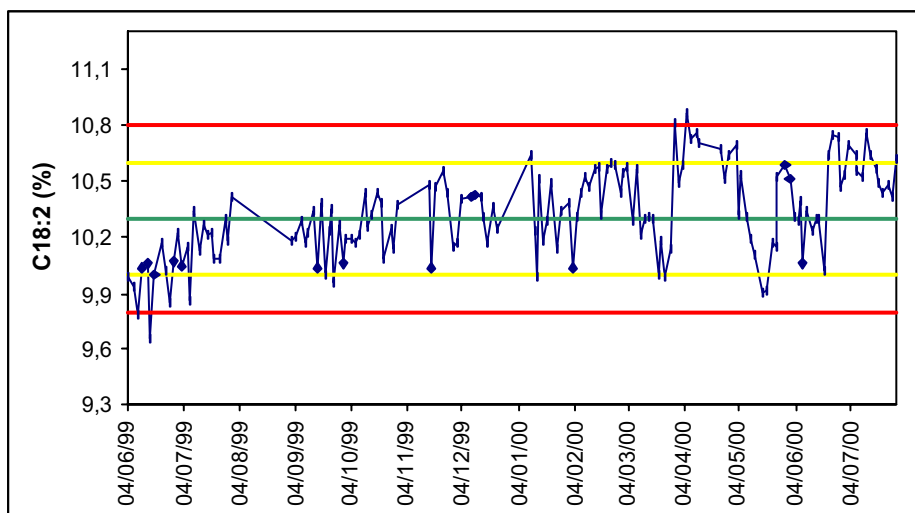


Figura 4.3.27. Gráfico de control de valores predichos para C18:2 en colectivo C.



Como puede apreciarse en dichas figuras, los valores predichos a lo largo de los meses de estudio para los diferentes ácidos grasos, presentan pequeñas oscilaciones aleatorias alrededor de los valores medios (líneas verdes horizontales). Los límites de aviso son superados en varias ocasiones en todos los ácidos grasos, lo cual indica que el sistema se aproximaba al máximo error de predicción admisible para las ecuaciones seleccionadas. Sin embargo los límites de actuación, los cuales informan de que las ecuaciones se encuentran fuera de control en cuanto a exactitud se refiere, únicamente se sobrepasaron en 3 determinaciones para los ácidos C18:0 y C18:2. Sin embargo, una vez sobrepasados dichos límites en cada una de las determinaciones anómalas, el

siguiente valor predicho se encontraba de nuevo dentro de los límites de actuación, con lo cual no fue necesario realizar ningún tipo de acción para corregir la presencia de errores sistemáticos.

Estos resultados muestran la alta exactitud y precisión y robustez de las ecuaciones seleccionadas al poder predecir de forma exacta y precisa la composición en ácidos grasos mayoritarios en una muestra de referencia de grasa de cerdo Ibérico durante un periodo de 13 meses.

4.3.1.7 Evaluación del pretratamiento de derivadas sobre al cálculo del estadístico H

Con independencia de los resultados obtenidos a nivel de repetibilidad del valor predicho de una determinada ecuación de calibración NIRS, la aplicación en rutina de dicha ecuación debería implicar el utilizar algún algoritmo basado en la medida de la distancia espectral de la muestra a analizar al espectro medio del colectivo de calibración.

Como se ha indicado en el apartado 3.3.1.4, a fin de simplificar el cálculo de distancias multivariantes, es recomendable el uso de métodos de síntesis de la información espectroscópica, previo al cálculo de dichas distancias, tales como el análisis de componentes principales (ACP). Asimismo se ha indicado la generalización del uso de la distancia de Mahalanobis, como estadístico que informa de la similitud o divergencia de un espectro con respecto al espectro medio del colectivo de calibración, lo que proporciona un método objetivo que evita un uso inadecuado del modelo predictivo o la realización de extrapolaciones de dicho modelo (Shenk *et al.*, 1991a,b, Shenk *et al.*, 1992; Shenk y Westerhaus, 1995b; ASTM, 1994 y 2000; Næs *et al.*, 2002).

Generalmente, los programas quimiométricos de tratamiento de datos NIRS, permiten el cálculo de la distancia de Mahalanobis, o de algún otro estadístico similar, tal como el denominado *leverage* incluido en el programa Unscrambler (Martens y Næs, 1989; Esbensen *et al.*, 1994;) o el estadístico H en el programa ISI (Shenk y Westerhaus, 1995 a,b). En particular, el paquete quimiométrico ISI presenta la ventaja de ser muy útil para la detección de anómalos durante el análisis de rutina, dado que realiza el cálculo del estadístico H justo después de obtener su espectro NIR y

paralelamente a la obtención de los datos predichos y al estadístico T de Student, en nuestro caso correspondientes a la predicción del contenido en ácidos grasos.

Conviene indicar, que el estadístico H utilizado por el programa ISI, fue definido para colectivos espectroscópicos correspondientes a muestras de forrajes, granos y otros alimentos para animales de bajo contenido en humedad (Shenk y Westerhaus, 1991a y b). Dichos autores señalan que aquellas muestras que presenten un valor de H superior a 3 deberían considerarse como anómalas espectrales. A lo largo de nuestro trabajo de evaluación y aplicación en rutina de ecuaciones NIRS para la predicción el contenido en ácidos grasos, se ha observado que el comportamiento de dicho estadístico, para el caso de datos espectrales NIR de grasa fundida de cerdo Ibérico, no parecía ajustarse a dicha regla. Por ello, se decidió profundizar en este aspecto, tratando de entender e interpretar los valores de H obtenidos para las muestras del colectivo de calibración.

Esta interpretación resulta esencial no solo desde el punto de vista científico sino también desde un punto de vista económico de aplicación de las ecuaciones desarrolladas ya que, en un principio, aquellas muestras que presenten un valor del estadístico H superior al límite máximo recomendado ($H=3$), deberían ser enviadas al laboratorio para ser analizadas por el método de referencia (Shenk *et al.*, 1992), en nuestro caso, por cromatografía gaseosa. Dado que el coste del análisis mediante los métodos de referencia, conlleva un coste económico y de tiempo, es imprescindible el contar con un método automatizado, que permita poder diferenciar con exactitud, las muestras con datos espectrales anómalos verdaderos de aquellas que no lo sean.

Para profundizar en el estudio del estadístico H se emplearán los valores predichos de este estadístico para los colectivos de validación A, B y C, obtenidos en períodos diferentes al de la obtención de los espectros del colectivo de calibración. Así, como se indicó en el apartado 3.3.1.3, los espectros de las muestras pertenecientes al colectivo de calibración se obtuvieron entre enero de 1.997 y abril de 1.998, los pertenecientes a las muestras del colectivo de validación A se tomaron justo tras analizar el colectivo de calibración (mayo de 1.998) y los incluidos en el colectivo de validación B, 19 meses después (noviembre de 1.999). Los diferentes espectros de la muestra del colectivo de validación C se obtuvieron de forma periódica lo largo de 13

meses (de junio de 1.999 a julio de 2.000), un año después de analizar por NIRS las muestras del colectivo de calibración.

Como se ha mostrado anteriormente, los diferentes tratamientos de derivación parecen corregir o minimizar fuentes de variación no controladas (instrumentales, operacionales y/o ambientales) que afectan al error de predicción (Tabla 4.3.17). Sin embargo, los resultados obtenidos en el apartado 4.3.1.4, fueron contradictorios en cuanto al valor medio del estadístico H para cada colectivo de validación.

El valor medio del estadístico H para el colectivo de validación A fue 2,63, inferior por tanto el valor máximo recomendado ($H=3$). Esto significaría que los espectros del dicho colectivo son similares a los espectros del colectivo de calibración. Sin embargo, los valores medio del estadístico H para los colectivos de validación B ($H=25,44$) y C ($H=33,38$) fueron muy superiores al valor máximo recomendado. Por este motivo, las muestras de dichos colectivos B y C habrían de considerarse como muy diferentes a las muestras incluidas en el colectivo de calibración y deberían de ser enviadas a analizar mediante cromatografía de gases, a pesar de los excelentes valores de ETP obtenidos en la predicción de las mismas.

En realidad, estos valores anómalos del estadístico H para los colectivos de validación B y C no tendrían sentido ya que el colectivo B contiene espectros de muestras que estaban incluidas en el colectivo de calibración mientras que el colectivo C contiene espectros de una única muestra similar en cuanto composición y valores de absorbancia a la composición y espectro medio del colectivo de calibración. Además, como ha sido indicado anteriormente (apartado 4.3.1.4), todos los estadísticos utilizados para la evaluación de las ecuaciones mediante el procedimiento de validación externa propuesto por Shenk y Westerhaus (1995b) al ser aplicado a los colectivos de validación B y C se encontraban dentro de los límites establecidos.

Shenk y Westerhaus (1995b) recomiendan un pretratamiento de derivadas del tipo (1,4,4,1) para la aplicación del algoritmo CENTER, el cual en definitiva, permite posteriormente el cálculo de distancias espectrales durante el análisis de rutina. Teniendo en cuenta la importancia ya mencionada de dicho estadístico se planteó,

además de evaluar su valor en diferentes colectivos de validación, el profundizar en su comportamiento dependiendo de la estrategia de derivación usada.

Por ello, se compararon los valores de H obtenidos a partir de la aplicación del programa CENTER incluido en el software ISI a los datos espectrales del colectivo de calibración inicial sometidos a diferentes tratamientos de derivadas. Como tratamientos de derivadas se emplearon aquellos detectados como óptimos en la selección final de las ecuaciones de calibración (apartado 4.3.1.6), es decir, (1,10,10,1), (2,5,5,1), (2,10,5,1) y (3,10,5,1).

La Tabla 4.3.21 recoge aquellas muestras cuyo espectro muestra valores de H superiores a 3 en los sucesivos análisis de componentes principales realizados a partir de los datos espectrales del colectivo de calibración sometido a los pretratamientos de derivación mencionados y sin tratamiento de derivación (0,0,1,1).

Tabla 4.3.21. Valores del estadístico H de las muestras anómalas, del colectivo de calibración, eliminadas tras los sucesivos análisis de componentes principales, para los diferentes tratamientos de derivadas.

Muestras anómalas	Tratamiento de derivadas				
	0,0,1,1	1,10,10,1	2,5,5,1	2,10,5,1	3,10,10,1
1	3,38**				
7	3,68**				
4	4,55*	4,81*	3,27*	4,61*	3,37*
5	4,19*	5,24*		4,73*	
28*				3,02***	
69	4,65*	3,44*	4,49*	4,57*	4,88*
69-L	3,03**		3,69*	3,92**	3,19*
70	4,57*	3,59*	4,27*	3,93*	4,16*
114	4,36*	3,81*	4,02*	3,98*	4,10*
116	11,66*	12,44*	13,11*	14,19*	13,76*
143			3,31**	3,45**	
98120	4,22*	4,81*	4,29*	3,35*	
98218	5,18*	4,58*		3,46*	3,04**
TOTAL	11	8	8	11	7

* muestra eliminada en el 1^{er} ACP; ** muestra eliminada en el 2^o ACP; *** muestra eliminada en el 3^{er} ACP.

Como se aprecia en dicha tabla, las muestras detectadas como anómalas son similares para todos los pretratamientos, si bien existe cierta disparidad en el número de muestras detectadas como anómalas, siendo de 11 para el tratamiento (0,0,1,1), 8 para (1,10,10,1), 8 para el (2,5,5,1), 11 para el (2,10,5,1) y 7 para el (3,10,10,1). Esto indicaría que el espacio vectorial definido en cada caso es ligeramente diferente, lo cual

se confirma al observar los valores de H calculados para una misma muestra en cada espacio.

Así, la muestra 98218 presentó una distancia H de 5,18 para un tratamiento (0,0,1,1), 4,58 para (1,10,10,1), 3,46 para (2,10,5,1), 3,04 para (3,10,10,1) e incluso resultó con una distancia inferior a 3 (2,70) para un tratamiento de derivada (2,5,5,1). Con ello se muestra la importancia del tratamiento de derivación en el cálculo de la distancia H y, por tanto, en la detección de anómalos espectrales utilizando como criterio el valor de 3 como límite para el estadístico H .

Para poder confirmar la mayor o menor influencia del tratamiento de derivación en el cálculo de la distancia H , y asimismo en la definición del valor límite de dicho estadístico para considerar muestras anómalas, es necesario realizar una mayor caracterización e interpretación de los espectros considerados como anómalos. Para ello se emplearon las anotaciones acerca de las incidencias detectadas durante la toma de espectros, los datos de composición química y la representación gráfica del colectivo espectral de calibración.

La muestra con mayor valor de distancia H para los diferentes tratamientos de derivadas considerados fue la 116. De acuerdo con las anotaciones en el cuaderno de laboratorio, la grasa de dicha muestra presentó un aspecto mucho más turbio al resto, debido posiblemente a la presencia de restos de tejido adiposo o conjuntivo. Ello podría dar lugar a que dicha muestra se diferenciara del resto de muestras del colectivo, factor que se vería reflejado en su espectro NIR. Dicha muestra fue considerada como anómala (valores de $H > 11$) en todos los tratamientos de derivadas considerados.

De acuerdo con la información incluida en el cuaderno de laboratorio, en la recogida de espectros de la muestra 98218 se observó la presencia de alguna burbuja en la cápsula. Teniendo en cuenta los resultados del ensayo realizado para la puesta a punto de una metodología de análisis de grasa fundida, la presencia de alguna burbuja es un factor que afecta a la repetibilidad espectroscópica NIR. Sin embargo, en el proceso de desarrollo de calibraciones, la validación cruzada permite observar unos valores predichos correctos para esta muestra, con valores de residuales par los cuatro ácidos grasos estudiados muy bajos. Es decir, la predicción que realiza el modelo a partir de su espectro es correcta. Sin embargo esta muestra fue detectada como anómala en 4 de los

5 tratamientos de derivadas empleados, si bien, en uno de ellos (3,10,10,1) lo fue tras un 2º pase de detección de anomalos y con un valor de H ($H=3,04$) muy próximo al valor límite considerado.

Todo ello provoca que no sea tan obvio el considerar dicho espectro como anómalo, a pesar de la presencia de algunas burbujas en la cápsula. De hecho, como se mostró en el apartado 4.2.1, la influencia de la presencia de burbujas en la repetibilidad de la medida espectral solo es significativa cuando las burbujas ocupan una parte importante de la superficie de ventana, hecho que no ocurrió en este caso.

Aunque no se ha mostrado gráficamente en este documento, la representación gráfica de los espectros del colectivo de calibración muestra claramente la presencia de unos espectros diferentes al resto. Estos son los correspondientes a las muestras 4, 69 y 70. En el caso de la muestra 4, su espectro es el que presenta los menores valores de absorbancias para todo el colectivo a lo largo de todo el rango de longitudes de onda. En el caso de las muestras 69 y 70, sus espectros poseen los máximos valores de absorbancia en todo el rango de longitudes de onda y se encuentran claramente diferenciados del resto de espectros del colectivo.

Las muestras 4 y 69 presentaron valores extremos de composición en ácidos grasos C16:0 (24,3 y 24,8% respectivamente), C18:0 (14,0 y 14,2% respectivamente) y C18:1 (48,2 y 45,8% respectivamente) si bien existen otras muestras, incluso con similares valores extremos (Figura 4.3.1) que han presentado valores de H inferiores a 3 en todos los tratamientos de derivadas considerados. Ello, junto con las ya comentadas diferencias gráficas espectrales, llevaría a emitir la hipótesis de que los espectros de dichas muestras, no difieren significativamente del resto en razón de sus datos químicos, si no mas bien debido, a una incorrecta preparación de muestra y/o recogida de datos espectroscópicos.

La muestra 70 posee unos valores de composición en ácidos grasos que pueden considerarse como bien representados en el colectivo de calibración. Por ello y teniendo en cuenta su anomalía gráfica se podría indicar que, al igual que para las muestras 4 y 69, la existencia de alguna fuente de error no controlada, durante la preparación de muestra y/o recogida de datos espectroscópicos, es la responsable de su anomalía espectral.

Las muestras 4, 69 y 70 fueron detectadas como anómalas, en el primer análisis de componentes principales, para todos los tratamientos de derivación. Los valores de H de dichas muestras oscilaron entre 3,27 y 4,88.

Las muestras 5, 69-L y 98120 también poseen para casi todos los ácidos grasos mayoritarios valores extremos, muy próximos a los máximos o los mínimos de los correspondientes intervalos de composición. Así, La muestra 5 presentó valores extremos para los ácidos grasos C16:0 (25,3%), C18:0 (14,8%), C18:1 (45,0%) y C18:2 (7,8%). Asimismo, la muestra 69-L presentó valores extremos para los ácidos grasos C16:0 (24,7%), C18:0 (14,2%) y C18:1 (45,6%) mientras que la muestra 98120 únicamente presentó valores extremos para los ácidos grasos C16:0 (18,6%) y C18:1 (56,4%).

Dado que no existen argumentos, que puedan apoyar lo inadecuado de su análisis NIRS, parece lógico que su detección como muestras anómalas sea debida a las características peculiares de dichas muestras, tanto en composición en ácidos grasos, como en otras características físico-químicas desconocidas. La muestra 5 fue detectada como anómala por 3 tratamientos de derivación, con valores de H entre 4,19 y 5,24, las muestras 69-L y 98120 lo fueron en 4 tratamientos, con valores de H que oscilaron entre 3,03 y 4,81.

Con respecto a las muestras 1 y 7, se ha de indicar que sus espectros son similares al espectro medio del colectivo de calibración y únicamente presentaron una composición anormalmente alta para el ácido graso C18:2 (13,0% para la muestra 1 y 12,8% para la muestra 7). Sin embargo, muestras con similar e incluso mayor contenido en dicho ácido graso presentaron valores de H inferiores a 3. Todo ello, junto con el hecho de que dichas muestras únicamente fueran detectadas como anómalas por aquel tratamiento en el que no se realiza ninguna derivación, (0,0,1,1) permitiría concluir sobre lo inadecuado de la utilización de los datos espectroscópicos brutos, como paso previo a la aplicación del algoritmo CENTER y, en definitiva, al cálculo de las distancias H .

Respecto a la muestra 114, la información disponible sobre su composición química y la representación gráfica, de su espectro frente al resto de muestras del colectivo de calibración no nos informa de anomalías. Sin embargo, durante el proceso

de desarrollo de calibraciones, dicha muestra presentó en todos los casos, valores de los residuales (laboratorio-predicho NIRS) significativamente elevados ($T > 3$), lo que asimismo corrobora que la muestra difiere del resto de muestras del colectivo de calibración. Dicha muestra fue detectada como anómala en el primer análisis de componentes principales de todos los tratamientos de derivación, con valores de H oscilando entre 3,81 y 4,36, por lo que definitivamente debe considerarse como muestra anómala, probablemente debido a errores cometidos durante la preparación de muestra y/o recogida de datos espectroscópicos.

En similar situación a la muestra 114 en cuanto a información previa se encontrarían las muestras 28* y 143. Sin embargo, dichas muestras no fueron detectadas como anómalas de forma sistemática mediante los diferentes tratamientos de derivadas considerados. Así, la muestra 143 fue detectada como anómala por dos tratamientos de derivación (2,5,5,1 y 2,10,5,1) con valores de H de 3,31 y 3,45. La muestra 28* únicamente fue detectada como anómala tras un tercer análisis de componentes principales por el tratamiento de derivadas (2,10,5,1), presentando un valor de H de 3,02. Todo ello hace plantear ciertas dudas con respecto a la consideración de estas muestras como verdaderas anómalas.

Después de la anterior discusión de anómalos y, tras la observación de los resultados descritos en la Tabla 4.3.21, es difícil concluir acerca del mejor pretratamiento de derivadas para detectar las verdaderas muestras anómalas del colectivo de calibración. Un hecho que ha quedado patente en el estudio de los datos anómalos ha sido la detección, en ocasiones errónea, de un mayor número de muestras como anómalas al considerar los espectros sin ningún tipo de derivación. Esto permitiría confirmar la necesidad de utilizar un tratamiento de derivadas previo, al cálculo del análisis de las componentes principales realizado por el algoritmo CENTER (Shenk y Westerhaus, 1995b).

Sin embargo, resulta más difícil poder decantarse por alguno de ellos. La discusión anteriormente realizada de los resultados obtenidos permitiría concluir que aquellos tratamientos que han logrado detectar un mayor número de verdaderas muestras anómalas serían los del tipo (1,10,10,1), (2,5,5,1) y (2,10,5,1). Sin embargo, se ha de tener en cuenta que el tratamiento (2,5,5,1) no detecta algunas muestras con claras anomalías como son las muestras 5 y 98218, mientras que el tratamiento (2,10,5,1)

incluye como anómalas algunas muestras como la 28* y 143 que, de acuerdo con la información disponible, no deberían de ser consideradas como tales.

Todo lo anteriormente expuesto hace sospechar que algunas de las discrepancias observadas podrían estar causadas por fuentes de variación ambientales y/o instrumentales que al parecer afectan a la medida de la distancia H . Pero asimismo podría pensarse que el límite de $H=3$ establecido por Shenk y Westerhaus (1991a,b) para considerar muestras anómalas pudiera no ser adecuado para espectros de grasa fundida. De hecho, muestras como la 1, 7, 28* y 143 no son verdaderas anómalas.

Los resultados de la Tabla 4.3.21 no permiten confirmar las dos hipótesis planteadas en un principio, por lo que se decidió continuar el estudio del comportamiento del estadístico H .

Así, y dado que en el proceso de desarrollo y evaluación de calibraciones NIRS, la distancia H se utiliza en el procedimiento de validación externa, se podría evaluar el comportamiento del estadístico H en colectivos de validación no incluidos en la calibración. Los valores medios del estadístico H para cada uno de los colectivos de validación, tras emplear para su cálculo los diferentes espacios vectoriales desarrollados con anterioridad, aparecen reflejados en la Tabla 4.3.22.

Tabla 4.3.22. Valores medios del estadístico H para los colectivos de validación tras emplear diferentes espacios de componentes principales.

Grupo Validación	H (0,0,1,1)	H (1,10,10,1)	H (2,5,5,1)	H (2,10,5,1)	H (3,10,10,1)
Colectivo A	2,63	2,32	1,19	1,35	1,22
Colectivo B	25,44	23,42	21,87	14,39	13,99
Colectivo C	33,38	33,65	19,91	11,71	11,91

Tras una primera observación de dicha tabla se observa que, los valores medios del estadístico H calculados a partir de los espacios de las componentes principales de los espectros sometidos a diferentes tratamientos matemáticos son diferentes. Este hecho es normal ya que para el cálculo de cada espacio de componentes principales se parte de valores de variables diferentes y de un colectivo de muestras distinto.

Además se observa que, excepto para el colectivo de validación A, los valores medios del estadístico H son muy superiores al valor máximo recomendado, por

lo que deberían de considerarse dichos espectros como anómalos. También se aprecia en dicha tabla que, para los colectivos B y C, los valores medios de H son inferiores al emplear los tratamientos de derivadas (2,10,5,1) y (3,10,10,1). Para el colectivo de validación A, además de estos tratamientos de derivadas, también fue menor el valor medio del estadístico H para el tratamiento (2,5,5,1).

El hecho de que el valor medio del estadístico H para el colectivo A fuera inferior a 3 puede deberse a que dichos espectros se obtuvieron justo a continuación en el tiempo de los espectros incluidos en el colectivo de calibración. De esta forma, las muestras de ambos colectivos tuvieron la misma procedencia y similares condiciones instrumentales y ambientales de análisis, hecho que no ocurrió con los colectivos B y C.

Así, el colectivo B contenía espectros de muestras incluidas en el colectivo de calibración pero analizadas de nuevo 19 meses después. Pese a ello, los valores medios de H para dicho colectivo fueron, para todos los espacios vectoriales empleados, superiores a 14. Este hecho parece mostrar la elevada influencia de las condiciones instrumentales y/o ambientales en el cálculo de la distancia H .

Para profundizar en la influencia del periodo de recogida de espectros en los valores alcanzados por el estadístico H , se obtuvieron los valores de dicho estadístico sobre los espectros que constituían el colectivo de validación C. De esta forma era posible observar el comportamiento de dicho estadístico a lo largo de un amplio periodo de tiempo. La Figura 4.3.28 muestra los valores medios de H , para los espectros de dicho colectivo, y para diferentes tratamientos de derivación de los datos espectroscópicos, realizados previamente al cálculo de distancias.

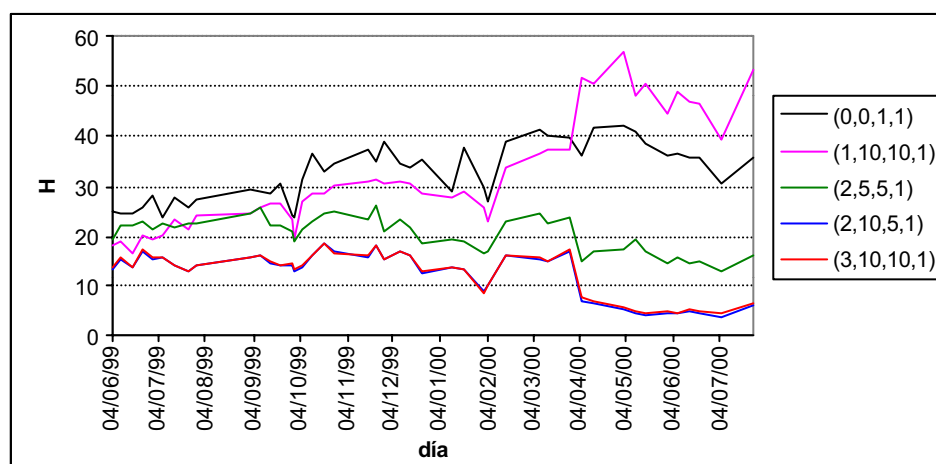
Un primer comentario acerca de los datos observados en dicha figura sería el de los valores tan elevados alcanzados por el estadístico H y sus enormes fluctuaciones dependiendo del tratamiento de derivadas considerado.

Como ya se apreciaba en la Tabla 4.3.22, en la Figura 4.3.28 todos los valores medios de H para el colectivo de validación C fueron superiores a 11 si bien, para algunos espectros las distancias H fueron inferiores, alcanzando valores cercanos a 5. Este menor valor del estadístico H se observó en algunas de las distancias calculadas a partir de espacios vectoriales con tratamientos de derivada de los espectros del tipo

(2,10,5,1) y (3,10,10,1) y únicamente para espectros obtenidos a partir del día 5 de junio de 2000.

Así, en esta fecha concreta, los valores de H calculados con dichos tratamientos de derivada para la muestra tomada con anterioridad (03/04/2000) fue 15 mientras que para el espectro obtenido posteriormente (07/04/2000) fue, aproximadamente, 7.

Figura 4.3.28. Valores medios de H para el colectivo C calculados a partir de diferentes espacios de componentes principales.



Este brusco descenso del valor de H coincidió en el tiempo con un severo ajuste de los valores de K y ϕ (apartado 4.3.1.5) dado que la exactitud en la medida de las longitudes de onda características del poliestireno superaba el valor límite recomendado (apartado 3.3.3.2). Esta coincidencia hizo sospechar acerca de la posible influencia de dicho factor de variación instrumental, sobre el cálculo del estadístico H en los colectivos de validación.

La Figura 4.3.29 muestra los valores de H para los mismos días en los cuales se determinó dicha precisión de la longitud de onda. Al mismo tiempo, la Figura 4.3.30 representa la evolución de la exactitud en la medida de las longitudes de onda (en el rango NIR) características del poliestireno y durante el periodo de análisis del colectivo C.

En ellas se aprecia una evolución, a lo largo del periodo considerado, de los valores de H para los tratamientos de derivada de (2,10,5,1) o (3,10,10,1) y de la exactitud en la medida de la longitud de onda durante el mismo periodo. Así, después de

una corrección de los valores K y ϕ , los valores de H presentan un descenso brusco desde valores aproximados de 15 a 7 dichos tratamientos de derivadas. La reducción de los valores de H al realizar la corrección mencionada anteriormente no se aprecia al no emplearse ningún pretratamiento de derivadas sobre los datos espectrales.

Figura 4.3.29. Valores medios de H para el colectivo C calculados con espacios de componentes principales obtenidos de espectros con tratamiento de derivación (2,10,5,1) y (3,10,10,1).

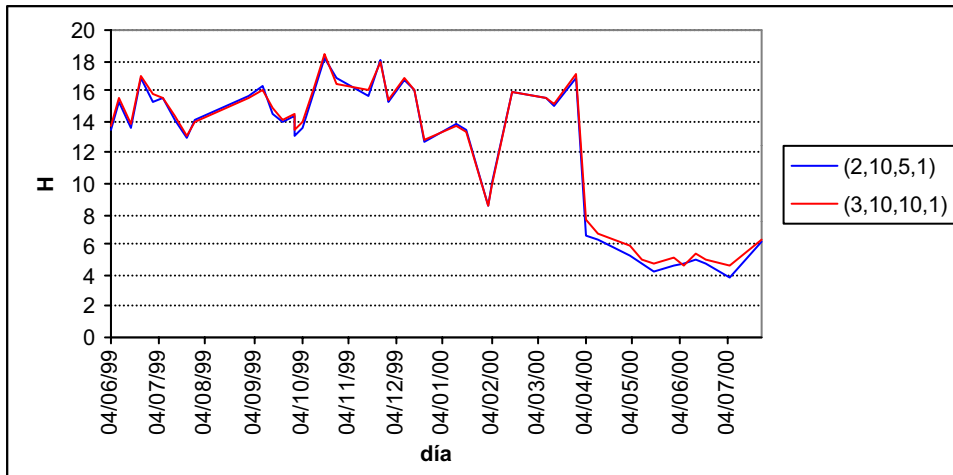
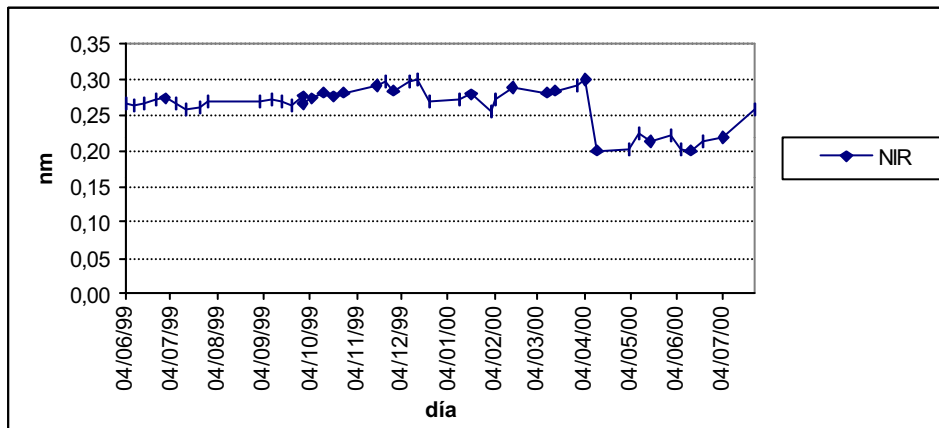


Figura 4.3.30. Evolución de la exactitud en la medida de las longitudes de onda (en el rango NIR) características del poliestireno y durante el periodo de análisis del colectivo C.



De acuerdo con estos resultados, se puede concluir sobre la gran importancia, en el análisis de grasas y aceites, de un control riguroso de la situación del equipo, evitando la toma de espectros cuando se aprecien cambios significativos en las condiciones de trabajo del instrumento (p.ej.: variaciones importantes en la ganancia de los detectores o en la precisión de la longitud de onda). Conviene llamar aquí la atención, sobre los resultados mostrados en las Figuras 4.3.12 a 4.3.15, correspondientes

al análisis de la de *Check Cell*. Dichos resultados muestran, que la muestra de control no mostró valores anómalos en la predicción de sus constituyentes en ninguno de los días en que se analizaron las muestras de los colectivos de validación considerados.

Por ello, antes del inicio de cada sesión de trabajo se recomienda la realización no solo de los controles recomendados por Shenk y Westerhaus(1995 a,b) y detallados en el apartado 3.3.3 sino también, la obtención de datos espectrales y predicho NIRS, de una muestra de grasa de referencia, de una composición química similar a la media del colectivo de calibración, tal como la utilizada en el colectivo de validación C.

Una vez conocida la influencia que sobre el estadístico H tienen las variaciones instrumentales y/o ambientales existentes día a día, se intentó modelizar estas variaciones, para lo cual, se creó un nuevo espacio n-dimensional utilizando los datos espectroscópicos del colectivo de calibración y los de la muestra de grasa de cerdo usada como referencia y analizada en todo el período de estudio. De esta forma, se utilizó en el cálculo de las componentes principales, no solo información espectroscópica explicada por la composición físico-química de las muestras (particularmente su composición en ácidos grasos), si no también la debida a la causada en el espectro por las variaciones debidas a factores ambientales y/o instrumentales existentes a lo largo del tiempo.

Para ello se aplicó de nuevo el algoritmo CENTER, empleando los pretratamientos de derivación reflejados en Tabla 4.3.21, si bien aplicados esta vez, sobre un nuevo colectivo formado por el colectivo de calibración inicial más el colectivo de validación C. La Tabla 4.3.23 recoge las muestras anómalas y sus correspondientes valores de H , obtenidos tras los sucesivos análisis de componentes principales realizados, para cada tratamiento de derivadas estudiado.

En este caso, se detectaron como anómalas un mayor número de muestras que para el colectivo de calibración anterior (Tabla 4.3.21), si bien, muchas de las muestras con valores de H superior a 3 del colectivo de calibración, continuaban detectándose como anómalas para este nuevo colectivo. Este fue el caso de las muestras 4, 5, 69, 69-L 114, 116, 98120 y 98218.

Sin embargo, también fueron detectadas como anómalas otras muestras procedentes del colectivo de validación C como las denominadas 17/12/99, 13/01/00, 24/01/00 y 09/02/00, cuyos espectros fueron recogidos en días donde la temperatura exterior y del laboratorio fue sensiblemente menor que durante el resto de días de dicho invierno. Como ya se indicó (apartado 4.2.1), la temperatura de la muestra es un factor que afecta sensiblemente a la repetibilidad espectral NIR, y ella podría ser la razón que explicara el que dichas muestras se detecten como anómalas a pesar de que dichos espectros proceden de una muestra cuyas características químicas se encuentran bien representadas en el colectivo de calibración.

Como sucedía con el colectivo anterior, el número de muestras detectadas como anómalas varía dependiendo del tipo de derivada aplicada sobre los espectros siendo 13 para un tratamiento (0,0,1,1), 8 para (1,10,10,1), 8 para (2,5,5,1), 10 para (2,10,5,1) y 9 para (3,10,10,1).

Tabla 4.3.23. Valores del estadístico H de las muestras anómalas, del colectivo de calibración y del colectivo C, eliminadas tras los sucesivos análisis de componentes principales, para los diferentes tratamientos de derivadas.

Muestras anómalas	Tratamiento de derivadas				
	0,0,1,1	1,10,10,1	2,5,5,1	2,10,5,1	3,10,10,1
17/12/99	3,11****		3,28**		3,04***
13/01/00	3,17***		3,85**	3,39***	3,08**
24/01/00	3,62*	3,11**	3,05*	3,80**	4,02**
09/02/00	3,13**				
1	3,10*				
4	3,38*	4,72*	3,20*	3,33*	3,19*
5	3,76*	3,86*	3,26*	3,77*	3,76*
69	4,00*			3,48**	3,04*
69-L		3,06***	3,76*	3,69**	3,08*
70	4,38*				
114	4,06*	3,73*	4,13*	3,61*	3,63*
116	11,67*	12,36*	13,76*	13,96*	14,22*
98218	5,37*	4,39*		3,23*	
98120	3,61*	3,86*		3,71*	
TOTAL	13	8	8	10	9

* muestra eliminada en el 1^{er} ACP; ** muestra eliminada en el 2^o ACP;

*** muestra eliminada en el 3^{er} ACP; **** muestra eliminada en el 4^o ACP.

Aquellas muestras consideradas como verdaderas anómalas en la anterior discusión de la Tabla 4.3.21 resultaron de nuevo seleccionadas al realizar los diferentes análisis de componentes principales del nuevo colectivo espectral. Este fue el caso de

las muestras 116, 4, 5, 69-L, 98120 y 114. Además, aquellas muestras sobre las que se plantearon ciertas dudas de ser consideradas como diferentes al colectivo considerado (1, 7, 28* y 143) no fueron detectadas como anómalas o bien lo fueron en el tratamiento que no usa derivadas (muestra 1). Con respecto a la muestra que presentaba algunas burbujas (muestra 98218), posee solo valores de $H > 3$ en el nuevo espacio n -dimensional construido, para el caso de no derivación y para las derivadas, (1,10,10,1) y (2,10,5,1).

Tras evaluar de forma conjunta los resultados obtenidos en las Tablas 4.3.21 y 4.3.23 y para el colectivo de grasa fundida considerado en este trabajo, los mejores resultados se han obtenido con el tratamiento de derivadas del tipo (2,10,5,1), el cual es aquel tratamiento que detecta todas las muestras consideradas como verdaderas anómalas. El hecho de que dicho tratamiento detecte como anómalas las muestras 13/01/00 y 24/01/00 podría deberse al efecto ya comentado de la temperatura.

Otra conclusión que es posible extraer de los datos de la Tabla 4.3.23 es que efectivamente, el valor de $H=3$ parece adecuado como límite para muestras de grasa fundida, una vez controladas las alteraciones en el valor de H causadas por diferentes fuentes de error.

Finalmente se procedió a calcular los valores de H para las muestras de los colectivos de validación A, B y C, estudiados anteriormente, utilizando la proyección de sus muestras sobre el espacio n -dimensional construido con los datos espectroscópicos del colectivo de calibración y del colectivo C unidos (Tabla 4.3.24).

Tabla 4.3.24. Valores medios del estadístico H para los colectivos de validación tras emplear diferentes espacios de componentes principales.

Grupo Validación	H (0,0,1,1)	H (1,10,10,1)	H (2,5,5,1)	H (2,10,5,1)	H (3,10,10,1)
Colectivo A	2,19	1,95	1,13	1,83	1,19
Colectivo B	2,10	2,01	2,12	2,16	2,28
Colectivo C	1,17	0,90	0,93	0,93	0,97

Como puede apreciarse, los valores del estadístico H para los 3 ficheros de validación y los diferentes tratamientos de derivadas fueron inferiores al valor máximo recomendado ($H=3$). Por lo tanto, este nuevo espacio, catalogaría a dichos colectivos como similares espectralmente al colectivo de calibración, tal como debería ser.

Los valores medios de H para el colectivo A fueron similares, para todos los tratamientos de derivadas, a los obtenidos a partir de los espacios de componentes principales definidos por el colectivo de calibración (Tabla 4.3.22). Sin embargo, para los colectivos B y C los valores medios del estadístico H disminuyeron de forma importante tras emplear los nuevos espacios de componentes principales, siendo inferiores a 2,28 para el colectivo B y 1,17 para el colectivo C, independientemente del tratamiento de derivadas empleado (Tabla 4.3.24).

Esta disminución del valor medio del estadístico H para el colectivo de validación C es lógica ya que él mismo ha participado en la propia definición del espacio de componentes principales. Sin embargo, esto no explicaría el descenso del valor de H para el colectivo de validación B. En este caso, la clara disminución del valor de H para los diferentes tratamientos de derivadas considerados, nos indicaría el que la estrategia utilizada verdaderamente ha posibilitado el modelizar las variaciones instrumentales y/o ambientales que afectaron al instrumento NIRS en ese periodo de análisis, ya que los espectros de dicho colectivo fueron tomados en una fecha incluida dentro del periodo de análisis del colectivo de validación.

Tras la realización del trabajo experimental de esta Tesis Doctoral y como continuación de la misma, en el Dpto. de Producción Animal de la Universidad de Córdoba se ha puesto a punto una metodología para incremento de la robustez, de las ecuaciones de ácidos grasos en grasa de cerdo Ibérico, haciendo a las mismas insensibles a variaciones instrumentales, ambientales u operacionales desconocidas. La metodología se basa en el uso en calibración de un fichero de repetibilidad constituido por un colectivo similar al colectivo C diseñado en esta Tesis Doctoral (Pérez-Marín *et al.*, 2002).

4.3.1.8 Obtención de ecuaciones para la determinación de ácidos grasos minoritarios

En los anteriores apartados se ha descrito el desarrollo de ecuaciones NIRS para la determinación de ácidos grasos mayoritarios (C16:0, C18:0, C18:1 y C18:2) en grasa fundida de cerdo Ibérico. Estos ácidos grasos constituyen aproximadamente el 95% del total de ácidos grasos presentes en esta grasa animal. El resto de ácidos saturados e insaturados (C14:0, C17:0, C17:1, C18:3, C20:0, C20:1), los cuales no

intervienen en la evaluación analítica de la alimentación en el contrato tipo homologado, constituyen los ácidos grasos minoritarios en esta grasa animal, sumando todos ellos el 5% restante.

En este estudio se han desarrollado ecuaciones NIRS para la determinación de cada uno de estos ácidos grasos minoritarios. La predicción rápida e instantánea de todo el perfil de ácidos grasos podría ser de utilidad para la puesta a punto de un sistema de control de calidad basado en todos los ácidos grasos presentes en la grasa del cerdo Ibérico. De hecho y como se ha indicado en el apartado 2.1.3, De Pedro (1995 y 2000) propone la implantación de herramientas de evaluación del régimen alimenticio de cerdo Ibéricos mediante el uso del perfil completo de composición en ácidos grasos y métodos de análisis multivariante.

La Tabla 4.3.25 muestra la composición media, los valores máximo y mínimo y la desviación típica para cada ácido graso minoritario del colectivo de calibración mientras que en la Figura 4.3.31 se muestran sus diagramas de distribución de frecuencias de las muestras a lo largo del intervalo de composición .

Tabla 4.3.25. Composición en ácidos grasos minoritarios de las muestras empleadas como colectivo de calibración (n=341).

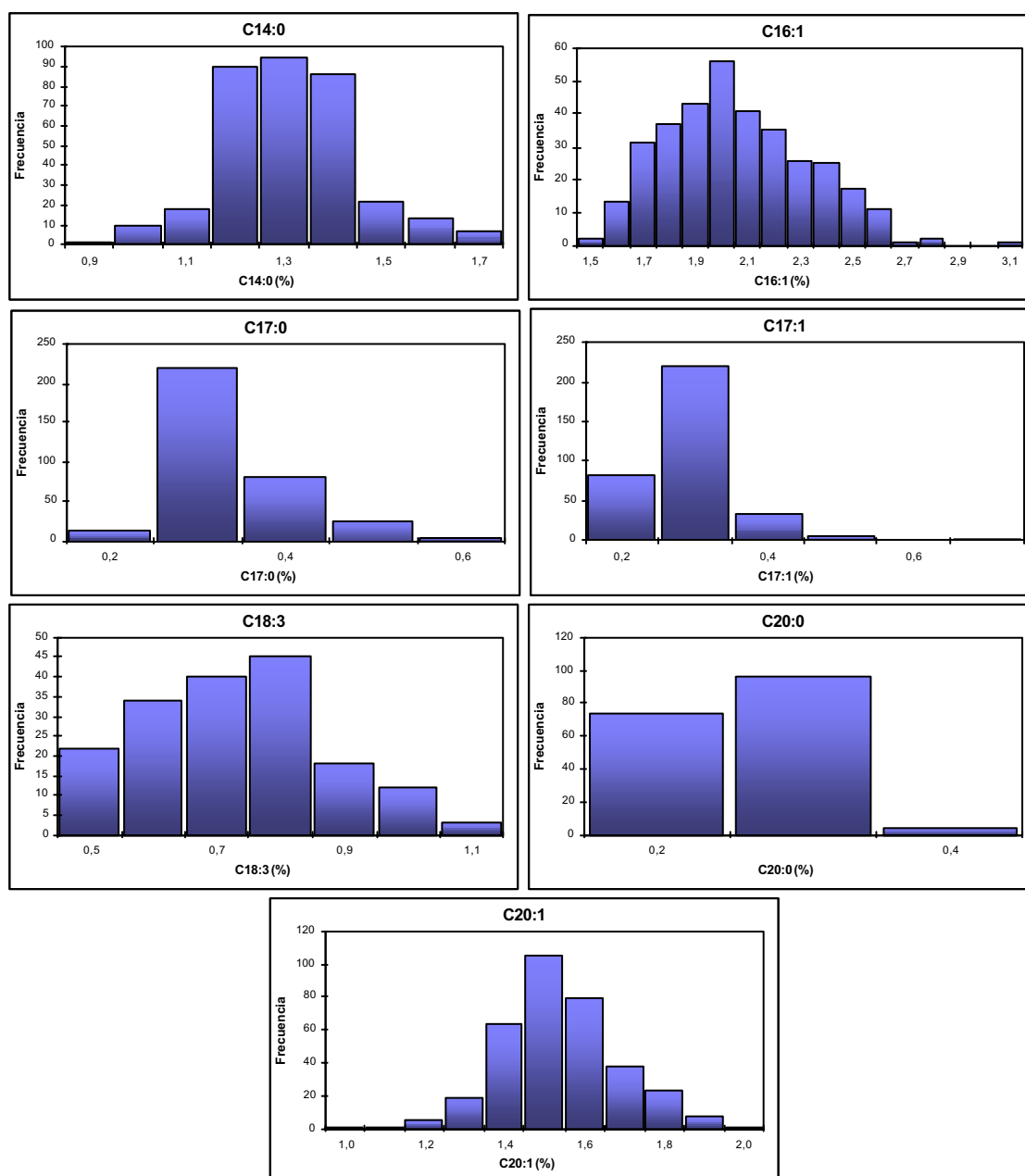
Ácido graso	Media	DT	Mínimo	Máximo
C14:0 (%)	1,31	0,14	0,90	1,70
C16:1 (%)	2,06	0,27	1,50	3,10
C17:0 (%)	0,34	0,07	0,20	0,60
C17:1 (%)	0,29	0,07	0,20	0,70
C18:3 (%)	0,73	0,15	0,50	1,10
C20:0 (%)	0,26	0,05	0,20	0,40
C20:1 (%)	1,54	0,15	1,00	2,00

Los datos recogidos en la Tabla 4.3.25 muestran que cada uno de los ácidos grasos se encuentra en la grasa de cerdo Ibérico a una concentración inferior al 3%, oscilando entre una composición media del 2,06% para el ácido C16:0 y del 0,26% para el ácido C20:0. Además, los intervalos de variación y los valores de desviaciones típicas son muy inferiores a los existentes en los ácidos grasos mayoritarios (Tabla 4.3.1).

Tras observar los datos incluidos en la Tabla 4.3.25 y la Figura 4.3.31, se observa que las muestra del colectivo de calibración cubren adecuadamente el intervalo de variabilidad existente en los ácidos grasos minoritarios para la grasa de cerdo Ibérico.

Se ha de destacar que los ácidos grasos presentes en menor concentración (C17:0, C17:1 y C20:0) poseen valores de composición prácticamente constantes, siendo muy pequeños los intervalos de variación y sus valores de desviaciones típicas inferiores a 0,07.

Figura 4.3.31. Diagramas de distribución de frecuencias del contenido en ácidos grasos minoritarios del colectivo de calibración.



Con los datos espectrales y químicos de las muestras de este colectivo se procedió a la obtención de diferentes ecuaciones de calibración para cada uno de los 7 ácidos grasos minoritarios considerados (C14:0, C17:0, C17:1, C18:3, C20:0, C20:1),

de acuerdo con el protocolo para el desarrollo de ecuaciones de calibración NIRS descrito en el apartado 3.3.1.4. Para ello, se hizo uso de nuevo de la secuencia automática AUTOCAL, descrita en el apartado 3.3.1.4. De entre todas las ecuaciones generadas se seleccionaron aquellas con mayor precisión de acuerdo con los estadísticos r^2 y *ETVC*.

La Tabla 4.3.26 muestra los estadísticos obtenidos para las ecuaciones seleccionadas como de mayor exactitud y precisión. Estas ecuaciones se obtuvieron a partir de un tratamiento de derivadas de (1,4,4,1) para los ácidos C17:1 y C20:1, (2,5,5,1) para C16:1, (2,10,5,1) para C17:0, (3,5,5,1) para C14:0 y C20:0 y (3,10,5,1) para el ácido graso C18:3. El número de términos RMCP seleccionados mediante validación cruzada para cada ecuación fue de 1 para C17:0, 2 para C14:0, 6 para C20:0, 7 para C17:1 y 9 para C16:1, C18:3 y C20:1. En dicha tabla se incluye asimismo el valor de *ETL* para cada una de los ácidos grasos, calculado a partir del análisis por duplicado de 20 muestras incluidas en el colectivo de calibración.

Tabla 4.3.26. Estadísticos de ecuaciones NIRS para ácidos grasos minoritarios obtenidas a partir del colectivo de calibración (n=341).

Ac. graso	Media	DT	ETL ⁽¹⁾	ETC	R ²	ETVC	r ²	RPD	RER
C14:0 (%)	1,31	0,11	0,07	0,08	0,39	0,09	0,36	1,25	7,49
C16:1 (%)	2,05	0,26	0,10	0,09	0,89	0,10	0,86	2,67	16,05
C17:0 (%)	0,33	0,05	0,06	0,04	0,18	0,04	0,17	1,10	6,58
C17:1 (%)	0,29	0,06	0,03	0,05	0,37	0,05	0,26	1,16	6,99
C18:3 (%)	0,73	0,15	0,03	0,06	0,84	0,08	0,75	1,99	11,95
C20:0 (%)	0,26	0,05	0,01	0,03	0,63	0,04	0,45	1,35	8,07
C20:1 (%)	1,55	0,14	0,07	0,09	0,58	0,10	0,48	1,38	8,29

1: calculado a partir del análisis por duplicado de 20 muestras incluidas en la calibración

Un primer comentario acerca de estos resultados sería el relativo a los más bajos valores de los coeficientes de determinación tanto para la etapa de calibración (R^2) como para la de validación cruzada (r^2), en relación a los obtenidos en las ecuaciones desarrolladas para los ácidos grasos mayoritarios (Tabla 4.3.4, Figura 4.3.8 y Tabla 4.3.17). A excepción de las ecuaciones obtenidas para los ácidos C16:1 y C18:3 los valores de r^2 no superaron el valor de 0,50, límite indicado por Shenk y Westerhaus (1996) para considerar a una ecuación como de buena exactitud y precisión. A pesar de ello, los valores obtenidos de *ETVC* para las distintas ecuaciones generadas son de una magnitud similar a los errores de laboratorio (*ETL*) del método de referencia, obtenido éste tras el análisis de 20 muestras por duplicado.

Esta discrepancia entre los resultados obtenidos a partir de los valores de r^2 y de $ETVC$ para las distintas ecuaciones tendría su explicación tras considerar los valores de DT para cada ácido graso. Así, en la mayoría de estos ácidos grasos, los valores de DT y ETL son del mismo orden lo cual implicaría que la ecuación NIRS estarían intentando modelar el propio error del método de referencia y no la variabilidad en cuanto a composición en las muestras. Por este motivo, los valores de coeficiente de determinación o porcentaje de variabilidad existente en el colectivo de calibración explicada por las ecuaciones NIRS han de ser bajos. En aquellos ácidos grasos donde los valores de DT son mucho más elevados a los de ETL (C16:1 y C18:3), los coeficientes de determinación (r^2) son superiores.

Esta similitud entre los valores de ETL y DT también afecta a los estadísticos RER y RPD . Así, en la mayoría de los casos, estos estadísticos no superan los valores mínimos recomendados (3 para el RPD y 10 para el RER) para las ecuaciones desarrolladas. Únicamente se superan el valor límite establecido para el RER en el caso de las ecuaciones desarrolladas para los ácidos grasos C16:1 y C18:3, aquellos donde las diferencias entre ETL y DT son elevadas.

Los bajos valores de r^2 obtenidos, en comparación con las ecuaciones descritas para los ácidos grasos mayoritarios (Tablas 4.3.4 y 4.3.17), también se podrían explicar si se tiene en cuenta la ecuación descrita por Mark y Workman (1991) según la cual, el valor máximo de este estadístico vendría definido en función de los valores de ETL y DT , de acuerdo a la siguiente expresión:

$$R^2_{max} = 1 - \left(\frac{ETL}{DT} \right)^2$$

Así, los valores máximo de R^2 que se podrían obtener a partir del colectivo de calibración serían de 0,60 para C14:0, 0,85 para C16:1, 0 para C17:0, 0,75 para C17:1, 0,96 para C18:3 y para C20:0 y 0,75 para 20:1. Así, si se comparan estos valores con los obtenidos para el estadístico R^2 en la Tabla 4.3.26, se puede apreciar que se alcanza o se aproxima en gran medida al límite teórico en las ecuaciones descritas para los ácidos grasos C16:1, C17:0, C17:1, C18:3 y C20:1. Por ello, la única estrategia a seguir para poder incrementar dicho valor sería el minimizar el error de laboratorio (minimizar el valor de ETL) habida cuenta de, al cubrir el colectivo de calibración la

práctica totalidad de la variación esperada para dichos ácidos grasos, no se podrá incrementar en demasía el valor de DT .

En el caso de los ácidos grasos C14:0 y C20:0, los valores de R^2 alcanzados son aproximadamente la mitad del valor máximo teórico que podría alcanzar en dicho colectivo. Por ello, algunas de las estrategias a realizar para lograr un incremento de dicho estadístico serían, por una parte la comentada anteriormente de minimización del error de laboratorio y, por otra el uso de otras técnicas de regresión que modelicen en mayor medida la composición química en base a la información espectroscópica NIR.

En comparación con otros estudios recogidos en el apartado 2.3.2 para la determinación de ácidos grasos minoritarios en grasa fundida o aceite, se ha de indicar que los resultados incluidos en la Tabla 4.3.26 presentaron una elevada exactitud y precisión.

Así, los valores de $ETVC$ para los diferentes ácidos grasos minoritarios fueron inferiores a los obtenidos por Pérez Vich *et al.* (1998) en aceite de girasol (para el ácido C16:1) y González (1997) en grasa de cerdo Ibérico (para los ácidos C14:0, C16:1, C18:3 y C20:0). Los valores de R^2 obtenidos en la presente Tesis para los ácidos grasos minoritarios fueron superiores a los descritos por González (1997) en grasa de cerdo Ibérico. En el caso de la ecuación descrita para el ácido C16:1 en el trabajo de Pérez Vich *et al.* (1998), el valor de r^2 fue superior al descrito en la Tabla 4.3.26 al presentar dicho ácido graso en el aceite de girasol un mayor valor medio e intervalo de variabilidad que en la grasa de cerdo Ibérico.

Las ecuaciones descritas por Pérez Marín *et al.* (2002) para la determinación de los ácidos C14:0 y C16:1 presentaron semejante exactitud y precisión a las descritas en la Tabla 4.3.26.

Sin embargo, los resultados de r^2 expuestos por Garrido *et al.* (2002) para la determinación de los ácidos grasos minoritarios C12:0, C14:0, C16:1, C17:0, C17:1, C20:0 y C22:0 fueron mayores a los descritos en este apartado. Ello se podría deber a la mayor variabilidad que presentaron estos ácidos grasos en el colectivo de calibración considerado (diferentes grasas de origen animal). Sin embargo, los valores de $ETVC$ fueron del mismo orden de magnitud.

4.3.2 MODALIDAD DE INTERACTANCIA-REFLECTANCIA

4.3.2.1 *Obtención de ecuaciones para la determinación del contenido en ácidos grasos en tejido adiposo*

Los resultados expuestos hasta este apartado en la presente Tesis Doctoral han mostrado que la tecnología NIRS, con sus ventajas frente a las de la cromatografía de gases, podría reemplazar a dicha técnica con un indudable ahorro en tiempo y coste de análisis. Sin embargo, las etapas previas a realizar para la obtención de las muestras de grasa fundida como la fusión de la muestra, el llenado y limpieza de cápsula, etc. podrían llegar a dificultar su implantación real a nivel de las industrias o laboratorios del sector.

Por ello, en este apartado se desarrollarán y evaluarán ecuaciones de calibración NIRS para la predicción del contenido en ácidos grasos de tejido adiposo de cerdo Ibérico empleando para ello una sonda de fibra óptica acoplada al instrumento NIRS. Las características de esta sonda y la metodología de análisis ha sido descrita y discutida en los apartados 3.3.2.3 y 4.2.2 respectivamente.

A partir de los datos de referencia (cromatografía gaseosa) y de los datos espectroscópicos correspondientes a muestras controladas en las campañas de sacrificio 97/98 a 99/00, se obtuvieron ecuaciones de calibración para cada uno de los ácidos grasos mayoritarios (C16:0, C18:0, C18:1 y C18:2).

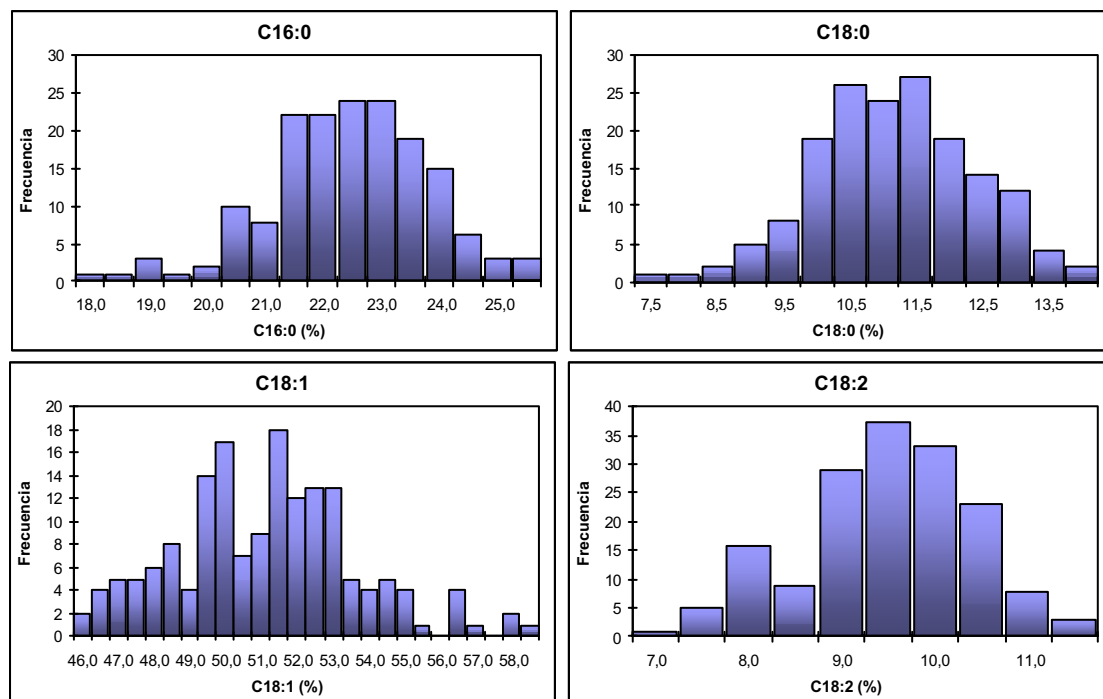
La Tabla 4.3.27 muestra la composición media, los valores máximo y mínimo y la desviación típica para cada uno de los ácidos grasos de la población de calibración.

Tabla 4.3.27. Composición en ácidos grasos de las muestras empleadas como colectivo de calibración (n=164).

Ácido graso	Media	DT	Mínimo	Máximo
C16:0 (%)	22,21	1,38	17,84	25,30
C18:0 (%)	10,94	1,21	7,21	13,90
C18:1 (%)	50,93	2,56	45,60	58,04
C18:2 (%)	9,27	0,92	7,00	11,40

Los datos recogidos en dicha tabla junto con los datos contenidos en los diagramas de frecuencias representados en la Figura 4.3.32, indican que la selección de muestras realizada permitió cubrir prácticamente todo el intervalo de variabilidad existente en los 4 ácidos grasos, para muestras de grasa de cerdo Ibérico sometidos a diferentes regímenes alimenticios (De Pedro, 1995 y 2000).

Figura 4.3.32. Diagramas de distribución de frecuencias del contenido en ácidos grasos en el colectivo de calibración global.



Una vez obtenidos los datos espectrales y químicos de las muestras del colectivo de calibración y de acuerdo con el protocolo para el desarrollo de ecuaciones de calibración NIRS ya descrito en el apartado 3.3.2.4, se procedió a la evaluación de anomalías espectrales mediante el cálculo del estadístico H de cada muestra al centroide de la población.

Tras realizar un primer análisis de componentes principales a partir de los datos espectrales considerando un pretratamiento de derivadas (1,4,4,1) de acuerdo con las recomendaciones de Shenk y Westerhaus (1995b), se obtuvieron 2 espectros con una distancia H superior al valor límite recomendado por el programa ($H=3$). Estos espectros pertenecieron a las muestras 990581 ($H=4,84$) y 980582 ($H=3,86$), no existiendo ningún motivo *a priori* para rechazarlas como anomalías si bien, estos altos valores de H hicieron sospechar de algún error en la recogida de los datos espectrales.

Una vez eliminadas estas muestras se realizó un segundo análisis de componentes principales obteniéndose únicamente 1 muestra anómala (07), siendo en este caso el valor de H ligeramente superior a 3 ($H=3,60$). Tras eliminar dicha muestra se repitió de nuevo el proceso, no presentando ninguna muestra distancias H al centro de la población superiores al valor recomendado como máximo. En total, se eliminaron 3 muestras en esta etapa de detección de anomalos espectrales, las cuales fueron eliminadas del colectivo de calibración.

Con los datos espectrales y químicos de las muestras de dicho colectivo se procedió a la obtención de diferentes ecuaciones de calibración para cada uno de los cuatro ácidos grasos considerados en este apartado, haciendo uso de la rutina AUTOCAL. De entre todas ellas se seleccionaron aquellas con mayor precisión de acuerdo con los estadísticos r^2 y $ETVC$.

La Tabla 4.3.28 muestra los estadísticos obtenidos para las ecuaciones seleccionadas como de mayor precisión. Estas ecuaciones se obtuvieron a partir de un tratamiento de derivadas de (2,10,5,1) para las variables C16:0 y C18:1 y de (0,0,1,1) para los ácidos grasos C18:0 y C18:2. El número de términos RMCP seleccionados mediante validación cruzada para cada ecuación fue de 6 para C16:0, 9 para C18:0, 8 para C18:1 y 11 para C18:2.

Tabla 4.3.28. Estadísticos de ecuaciones obtenidas a partir del colectivo de calibración (n=164).

Ac. graso	Media	DT	ETL ⁽¹⁾	ETC	R ²	ETVC	r ²	RPD	RER
C16:0 (%)	22,23	1,36	0,26	0,60	0,81	0,68	0,75	2,0	11,9
C18:0 (%)	10,96	1,15	0,22	0,41	0,87	0,47	0,84	2,5	14,8
C18:1 (%)	50,79	2,44	0,25	0,73	0,91	0,99	0,84	2,5	14,8
C18:2 (%)	9,34	0,87	0,15	0,34	0,85	0,40	0,78	2,2	12,9

1: estimado tras análisis mediante cromatografía de gases de 20 muestras por duplicado.

Un primer comentario acerca de estos resultados sería el del alto valor del coeficiente de determinación (superior a 0,75), tanto para el colectivo de calibración (R^2) como para el de validación cruzada (r^2) y para los cuatros ácidos grasos estudiados. De acuerdo con el criterio indicado por Shenk y Westerhaus (1996), estas ecuaciones han de ser consideradas como de alta precisión. Sin embargo, la precisión de estas ecuaciones es sensiblemente inferior a la de las desarrolladas para grasa fundida (Tabla 4.3.17). Los valores de $ETVC$ son superiores a los valores de ETL sí bien, para todas las

ecuaciones, son claramente inferiores a los valores de DT lo cual confirma la alta exactitud y precisión de estas ecuaciones.

Los valores de $ETVC$ superiores a los de ETL , en comparación con los resultados en grasa fundida donde ambos estadísticos eran del mismo orden, podrían deberse a numerosos factores. Entre ellos estaría él debido a las características inherentes de la muestra y el uso de la sonda de fibra que provocan un mayor error de repetibilidad espectral (apartados 4.2.1 y 4.2.2) como él debido a un posible error de muestreo a la hora de obtener un espectro representativo de toda la muestra de tejido adiposo.

Los valores del estadístico RER obtenido para todas las ecuaciones superaron el valor mínimo recomendado ($RER=10$). Sin embargo, los valores calculados de RPD fueron ligeramente inferiores al mínimo recomendado ($RPD=3$). Esta discrepancia entre los resultados de RER y RPD puede interpretarse teniendo en cuenta las fórmulas de dichos estadísticos y las características composicionales del colectivo de calibración. Así ambos estadísticos tienen similares fórmulas (apartado 3.3.1.4) si bien el valor de RPD se calcula a partir de la DT mientras que el valor de RER considera el valor del intervalo (máximo – mínimo). Como se aprecia en la Figura 4.3.32, los diagramas de frecuencias de la composición en ácidos grasos presentan unas curvas de distribución relativamente estrechas comparados con el amplio intervalo de composición que abarcan. Es decir, sus valores de DT son relativamente pequeños comparados con el valor de intervalo para cada ácido graso. Este hecho, junto con unos valores de $ETVC$ ligeramente altos, puede provocar la discrepancia observada entre los valores de los estadísticos RER y RPD . A pesar de ello, los resultados obtenidos confirman de nuevo la alta exactitud y precisión de las ecuaciones obtenidas y aseguran su validez desde el punto de vista de su aplicación a muestras no incluidas en la calibración.

Al comparar estas ecuaciones con las desarrolladas por Ripoche y Guillard (2001) para la determinación de los ácidos grasos C16:0 y C18:0 se puede apreciar la mejor exactitud y precisión de las ecuaciones descritas en la Tabla 4.3.28. Así, los valores de ETP en el trabajo de Ripoche y Guillard (2001) fueron muy superiores a los descritos en la presente Tesis mientras que los valores de r^2 fueron del mismo orden de magnitud.

En las Figuras 4.3.33 a 4.3.36 se exponen los valores de los coeficientes a distintas longitudes de onda de las ecuaciones descritas en la Tabla 4.3.28. Dichas figuras muestra que, al igual que para las ecuaciones obtenidas en grasa fundida, la mayor parte de las longitudes de onda para todos las ecuaciones, contribuyen con signo positivo o negativo en la predicción de cada ácido graso (coeficientes distintos de cero). Sin embargo, los valores de estos coeficientes son claramente menores a los obtenidos en las ecuaciones de grasa fundida, lo cual podría deberse a la diferencia en valores de absorbancia entre los espectros de tejido adiposo y grasa fundida.

Figura 4.3.33. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación de predicción para C16:0.

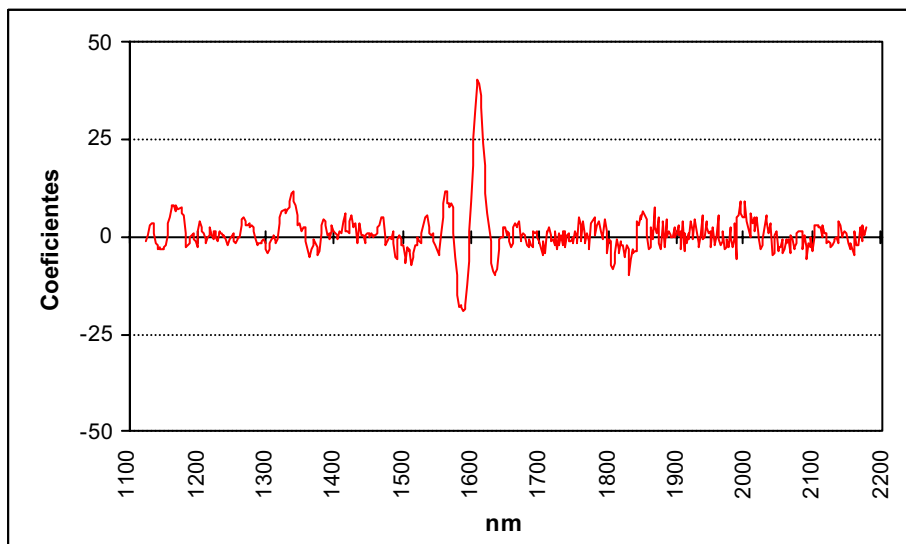


Figura 4.3.34. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación de predicción para C18:0.

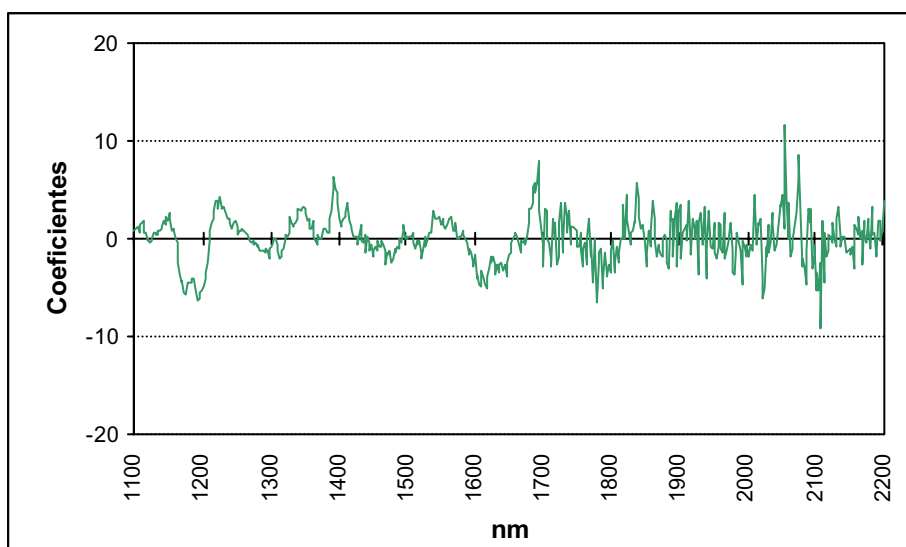


Figura 4.3.35. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación de predicción para C18:1.

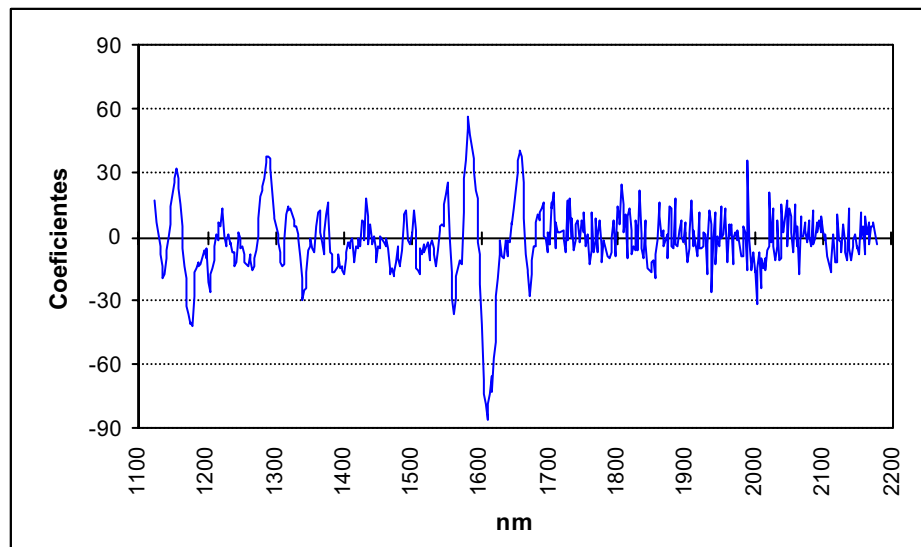
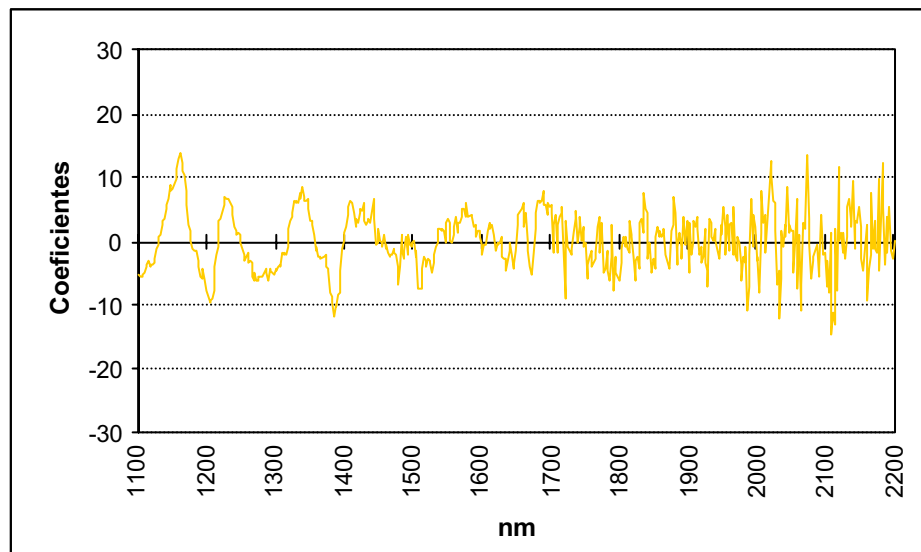


Figura 4.3.36. Valores de los coeficientes, a diferentes longitudes de onda, de la ecuación de predicción para C18:2.



Así, los valores de absorbancia para los espectros de tejido adiposo oscilan entre 0,5 y 1,7 unidades de absorbancia aproximadamente, apreciándose una clara pendiente positiva en la línea base espectral (Figura 4.2.8). En caso de los espectros de grasa fundida (Figura 4.2.1), los valores de absorbancias varían entre 0,2 y 1,1, teniendo en este caso una línea base claramente horizontal debido probablemente a una dispersión de la radiación (menor presencia del fenómeno de *scatter*). De esta forma, los valores de absorbancia de los espectros de tejido adiposo son, para la mayoría de las longitudes de onda, superiores a los valores de absorbancia espectral de grasa fundida. Este hecho provocaría que, para obtener una misma composición de un ácido graso, sea

necesario emplear coeficientes mayores en las ecuaciones de grasa fundida para obtener un valor similar en cada uno de los términos de la combinación lineal que define a una ecuación RMCP (coeficiente* Abs_{λ_i}).

Además, en las anteriores figuras se puede apreciar que no todas las longitudes de onda presentan la misma contribución en el valor final de composición predicha por la ecuación. Así, para los ácidos grasos C16:0 y C18:1, la región espectral comprendida entre 1.550 y 1.650 nm presenta una mayor aportación en los valores de ácidos grasos a predecir que el resto de longitudes de onda, aun a pesar de que en esta región no se aprecian bandas de absorción significativas de la grasa. En el caso de las ecuaciones para los ácidos C18:0 y C18:2, no se observa una diferencia clara entre los valores de los coeficientes en las diferentes regiones espectrales.

Un aspecto de estas gráficas que cabe resaltar es que los coeficientes obtenidos a partir de 1.700 nm presentan valores con una mayor aleatoriedad en las variaciones de los coeficientes. Esto puede deberse a la mayor importancia del ruido causado por la sonda de fibra óptica en la señal recibida por los detectores del instrumento, lo cual provocaría un descenso de la relación señal/ruido que afecta de forma negativa a la precisión de las ecuaciones desarrolladas.

Una vez obtenidas estas ecuaciones de calibración, se procedió a su evaluación mediante validación externa. Para ello se aplicaron las ecuaciones generadas sobre el colectivo de validación D, el cual contenía espectros de 15 muestras de grasa de cerdo Ibérico de características similares a las del colectivo de calibración pero no incluidas en él. Como se observa en la Tabla 4.3.29, la composición en ácidos grasos de este colectivo de validación es semejante, a nivel de valores medios, desviaciones típicas, máximos y mínimos, a la del colectivo utilizado para la calibración (Tabla 4.3.28).

Tabla 4.3.29. Composición en ácidos grasos del colectivo de validación D (n=15).

Ácido graso	Media	DT	Mínimo	Máximo
C16:0 (%)	21,93	1,94	19,28	25,60
C18:0 (%)	10,14	1,48	8,46	13,30
C18:1 (%)	51,72	3,13	46,50	56,63
C18:2 (%)	9,43	0,81	7,50	10,42

Los estadísticos obtenidos en el proceso de predicción de las muestras del colectivo D se indican en la Tabla 4.3.30. El valor de H medio en validación fue de 2,11, inferior al límite recomendado por el programa ($H=3$) requisito necesario, desde el punto de vista de la información espectral obtenida, para aplicar una ecuación NIRS a un colectivo distinto al de calibración.

Tabla 4.3.30. Estadísticos de validación externa del colectivo D (n=15).

Ac. graso	ETP	Sesgo	Límite sesgo	ETP(C)	Límite ETP(C)	Pendiente	R^2
C16:0 (%)	0,73	0,29	0,36	0,69	0,77	1,04	0,87
C18:0 (%)	0,52	-0,25	0,25	0,48	0,53	0,99	0,90
C18:1 (%)	1,03	-0,16	0,44	1,05	0,95	0,99	0,89
C18:2 (%)	0,55	0,04	0,20	0,57	0,44	0,75	0,57

Un análisis de los datos de la Tabla 4.3.30, teniendo en cuenta los criterios especificados en el apartado 3.3.1.4, confirma la alta exactitud y precisión en las predicciones realizadas. Así, los valores de ETP o error típico de los residuales (CG versus NIRS) para el colectivo evaluado y para cada ácido graso fueron similares a los valores de $ETVC$ (Tabla 4.3.28). El $sesgo$ obtenido para los diferentes ácidos grasos no superó el valor límite establecido para cada ecuación. Además, para el $ETP(C)$, sólo se superaron ligeramente los límites de dicho estadístico en las ecuaciones en las determinaciones de C18:1 y C18:2, siendo inferiores a los límites aceptados en las ecuaciones de C16:0 y C18:0.

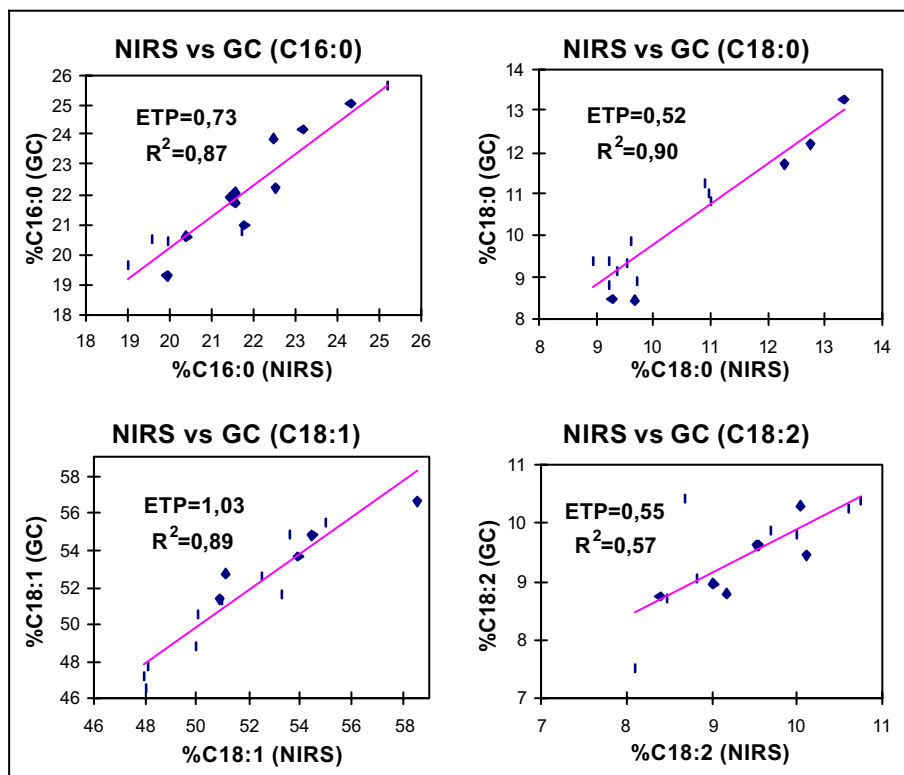
En los resultados de los estadísticos pendiente y R^2 se observó una discrepancia entre los resultados obtenidos para las diferentes ecuaciones desarrolladas. Así, para los ácidos C16:0, C18:0 y C18:1, los valores de pendiente se aproximaron al valor ideal de 1 mientras que los estadísticos R^2 fueron superiores a 0,87. Sin embargo, en las predicciones NIRS obtenidas para el ácido C18:2, los valores de pendiente y R^2 fueron claramente inferiores a 1. Este hecho puede explicarse tras observar de forma detallada las predicciones individuales de las muestras del colectivo de validación. En la predicción del ácido C18:2 en las muestras de este colectivo destaca la muestra 990501, la cual presentó un alto valor de residual (1,73) y por tanto, con una inadecuada predicción NIRS.

A pesar de ello, ni el valor de H de esta muestra ($H=1,73$) ni la información a priori que se tenía de ella podían hacer sospechar de esta inadecuada predicción. Si

esta muestra fuera eliminada del colectivo de validación, los estadísticos de predicción para el ácido C18:2 mejorarían sensiblemente, en particular la pendiente y R^2 . Así, el nuevo valor de *ETP* tras la eliminación de esta muestra sería de 0,34, mientras que la pendiente y R^2 se incrementarían a 0,88 y 0,83 respectivamente, unos valores claramente superiores a los obtenidos en un principio y similares a los estadísticos del resto de predicciones de ácidos grasos. Por todo ello, la causa más probable de la inadecuada predicción de esta muestra no se encuentre en el análisis ni predicción NIRS, sino en el valor de composición aportado por el método de referencia (cromatografía de gases).

La Figura 4.3.37 muestra las rectas de regresión de los datos predichos por NIRS versus los datos obtenidos por cromatografía de gases, donde se aprecia la excelente correlación entre ambos métodos de análisis y se confirman los resultados expuestos previamente. En la recta de regresión obtenida para el ácido C18:2, se observa con claridad la muestra 990501, dado su alto residual y, por tanto, su elevada distancia a la recta de pendiente 1. Este hecho provoca una clara disminución del valor de la pendiente y del estadístico R^2 .

Figura 4.3.37. Resultados de validación al emplear ecuaciones NIRS generadas mediante RMCPM.



4.3.2.2 Comparación de resultados mediante diferentes métodos de regresión: RMCPM vs. RLM

A partir del mismo colectivo de calibración empleado en el apartado anterior, se generaron ecuaciones de calibración NIRS para los ácidos grasos mayoritarios mediante el método de regresión lineal múltiple (RLM). En este caso, para el desarrollo de ecuaciones RLM en grasa fundida, únicamente se generaron ecuaciones NIRS para un máximo de 10 longitudes de onda seleccionadas por variable. En esta etapa se empleó de nuevo la secuencia automática AUTOCAL para la obtención de ecuaciones mediante RLM con diferentes tratamientos de derivadas (apartado 3.3.2.4).

La Tabla 4.3.31 muestra los estadísticos de las ecuaciones con mayor exactitud y precisión generadas mediante RLM a partir de una selección máxima de 10 longitudes de onda por variable. Como puede apreciarse, los valores de R^2 obtenidos durante el proceso de calibración fueron superiores a 0,78 y los *ETC* oscilaron entre 0,39% para C18:2 y 0,84% para C18:1. El tipo de derivada empleada en las ecuaciones con menor error varió entre una primera derivada para los ácidos C16:0 y C18:2 y la segunda derivada para las variables C18:0 y C18:1.

Tabla 4.3.31. Estadísticos de ecuaciones de calibración generadas mediante RLM, empleando 10 longitudes de onda por variable.

Ac. graso	N	Media	DT	Derivada	ETC	R^2
C16:0 (%)	159	22,24	1,36	1,10,10,1	0,64	0,78
C18:0 (%)	151	10,96	1,17	2,10,10,1	0,41	0,88
C18:1 (%)	157	50,87	2,53	2,5,5,1	0,84	0,89
C18:2 (%)	158	9,28	0,91	1,10,10,1	0,39	0,81

Como puede observarse tras la comparación de estos valores con los indicados en la Tabla 4.3.28, los estadísticos de calibración *ETC* y R^2 obtenidos mediante regresión RLM son similares a los logrados en las ecuaciones NIRS desarrolladas empleando una regresión RMCPM. Debido a la imposibilidad de poder realizar en el paquete estadístico ISI la validación cruzada durante la obtención de ecuaciones NIRS mediante regresión RLM, no se pudieron estimar ni por tanto comparar los estadísticos correspondientes.

La Tabla 4.3.32 muestra las longitudes de onda seleccionadas, sobre un máximo de 10, para cada una de las variables en las ecuaciones NIRS generadas

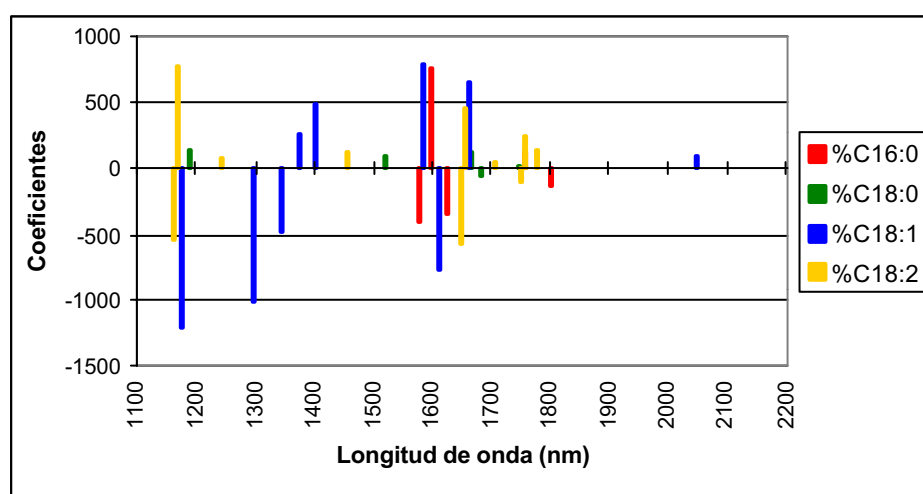
mediante RLM reseñadas anteriormente. En el caso del ácido C18:2, se seleccionaron el número máximo de 10 longitudes de onda. Sin embargo, para el resto de variables, el número de longitudes de onda seleccionadas por el algoritmo de regresión fue inferior a 10 (9 para C18:1 y 5 para C16:0 y C18:0).

Tabla 4.3.32. Longitudes de ondas seleccionadas para ecuaciones generadas mediante RLM.

Ac. graso	Longitudes de onda seleccionadas (nm)
C16:0	1576, 1598, 1624, 1758, 1800
C18:0	1190, 1520, 1666, 1680, 1746
C18:1	1172, 1292, 1338, 1370, 1398, 1582, 1608, 1658, 2046
C18:2	1160, 1166, 1242, 1452, 1646, 1654, 1704, 1748, 1754, 1774

Como puede apreciarse en dicha tabla y en la Figura 4.3.38, la mayor parte de las longitudes de onda significativas para todos los ácidos grasos, tras realizar los diferentes tests F, pertenecen a los segmentos de longitudes de onda comprendidos entre 1.150 a 1.200, 1.300 a 1.400 y 1.575 a 1.800 nanómetros. Además, cada ecuación selecciona longitudes de onda para cada variable dentro de todos los intervalos, no ciñéndose a una región en concreto. Esto indica que no se detecta ninguna región específica para la determinación de cada ácido graso sino que existen 3 regiones espectrales que contribuyen a dicha determinación.

Figura 4.3.38. Coeficientes de longitudes de onda seleccionadas en ecuaciones mediante RLM con un máximo de 10 longitudes de onda por variable.



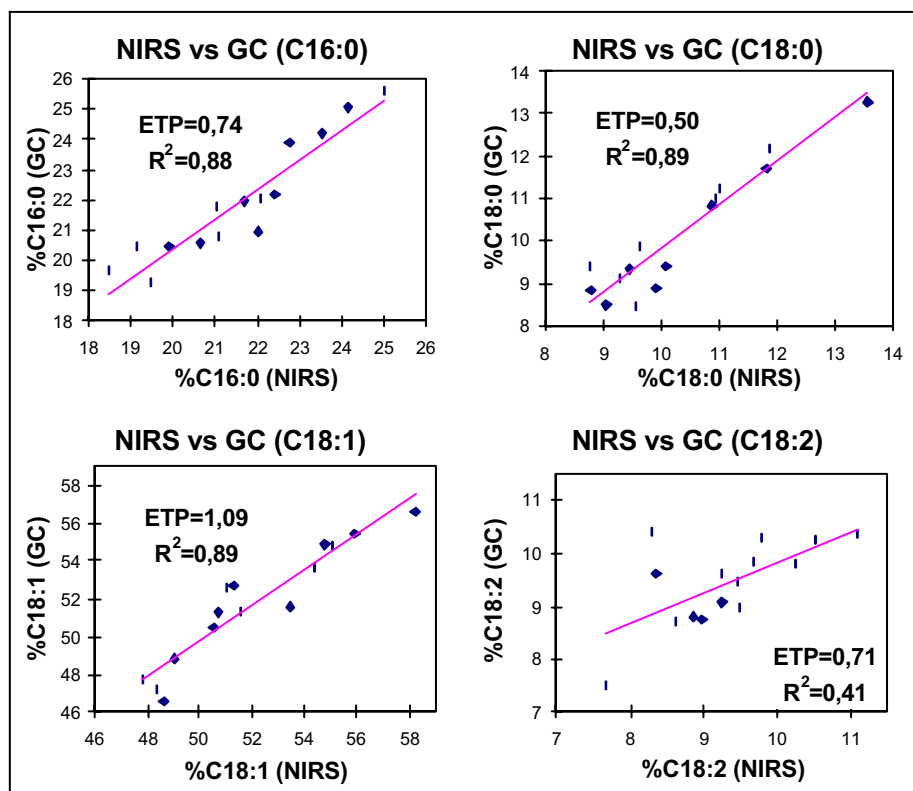
Al mismo tiempo, es importante destacar que dichos intervalos de longitudes de onda seleccionados mediante regresión RLM coinciden con las

principales bandas de absorción NIR de la grasa (no de tejido conectivo o agua) existente en el tejido adiposo de cerdo Ibérico.

Sin embargo, la contribución de cada longitud de onda en la determinación de un ácido graso es diferente, como se aprecia en la Figura 4.3.38. Así, para el ácido graso C18:1, la contribución de la longitud de onda a 2.046 nm es mucho más pequeña, evaluada desde el punto de vista de los coeficientes de la ecuación, que la contribución de la longitud de onda a 1.172 nm.

Una vez desarrolladas estas ecuaciones de calibración mediante regresión RLM, se procedió a su validación externa empleando para ello el colectivo de validación D (Tabla 4.3.29). La Figura 4.3.39 muestra, para cada ácido graso, las representaciones de los valores obtenidos mediante el método de referencia (cromatografía de gases) frente a valores predichos por NIRS a partir de las ecuaciones generadas mediante regresión RLM empleando un máximo de 10 longitudes de onda por variable (Tabla 4.3.32).

Figura 4.3.39. Resultados de Validación al emplear ecuaciones NIRS generadas mediante RLM, empleando 10 longitudes de onda por variable.



Como se puede apreciar, independientemente del ácido graso considerado, la representación de los valores se aproxima a una recta de pendiente 1, a excepción del ácido C18:2 donde dicha pendiente es menor.

Ello se confirma al evaluar los valores de R^2 y ETP obtenidos tras la predicción del contenido en ácidos grasos mayoritarios. (Figura 4.3.39). Así, para los ácidos C16:0, C18:0 y C18:1, los valores de R^2 fueron superiores en todos los casos a 0,88 y los valores de ETP oscilaron entre 0,50% para C18:0 y 1,09% para C18:1, del mismo orden a los valores de ETC obtenidos en las ecuaciones.

En el caso del ácido C18:2, los estadísticos presentaron una menor precisión, con un valor de ETP (0,71%) superior al ETC obtenido en la ecuación (0,39%) y un valor de R^2 excesivamente bajo. Este hecho se puede explicar de nuevo por la de nuevo en la excesiva diferencia detectada entre el valor obtenido por cromatografía de gases y la predicción NIRS para la muestra 990501 (residual=2,12).

Al eliminar esta muestra del colectivo de validación, los estadísticos de predicción para el ácido C18:2 mejoran sensiblemente, obteniéndose un nuevo valor de ETP de 0,48%, similar al valor de ETC de la ecuación, mientras que el R^2 se incrementaría a 0,69.

Indicar que tanto los valores de R^2 como de ETP para los 4 ácidos grasos mayoritarios fueron similares a los obtenidos tras la validación con las ecuaciones generadas mediante regresión RMCPM, por lo que dichos resultados de validación son independientes del método de regresión empleada en el desarrollo de ecuaciones.

4.3.2.3 Comparación de resultados mediante diferentes tratamientos de derivadas

En este apartado, al igual que se realizó para grasa fundida, se pretende estudiar la influencia del tratamiento de derivadas sobre la precisión de diferentes ecuaciones NIRS de predicción del contenido en ácidos grasos en tejido adiposo, generadas mediante regresión RMCPM. Asimismo, también se evaluará dicha influencia sobre los diferentes estadísticos de validación externa.

Para ello se generaron ecuaciones de calibración NIRS empleando el mismo colectivo de calibración descrito en el apartado 4.3.2.1 y el método de regresión RMCPM. Los diferentes tratamiento de derivadas, expuestos ya de forma detallada en el apartado 3.3.1.4, consistieron en 3 tratamientos diferentes de 1ª, 2ª y 3ª derivada y el correspondiente tratamiento de espectros sin derivar.

Al igual que en el caso de la grasa fundida, se empleó la secuencia automática AUTOCAL para el desarrollo de las ecuaciones de calibración teniendo en cuenta que se aplicó una corrección del fenómeno de radiación difusa del tipo *SNV+DTR*, se empleó el rango espectral entre 1.100 y 2.200 nm, se consideraron como límite de los estadísticos *T* y *H* los valores de 2,5 y 10 respectivamente y se fijó un número máximo de 10 pases de eliminación de muestras anómalas durante el desarrollo de las diferentes ecuaciones de calibración.

En las Figuras 4.3.40 y 4.3.41 se representan los valores de *ETVC* y r^2 respectivamente de las 10 ecuaciones de calibración para cada ácido graso considerado, generadas con diferente tratamiento de derivadas, empleando para ello la rutina de cálculo AUTOCAL.

Como se observa, la exactitud de las ecuaciones obtenidas para todas las variables puede considerarse como alta dados los valores de *ETVC* y r^2 .

Figura 4.3.40. Valores de ETVC obtenidos para ecuaciones desarrolladas con diferente tratamiento de derivadas.

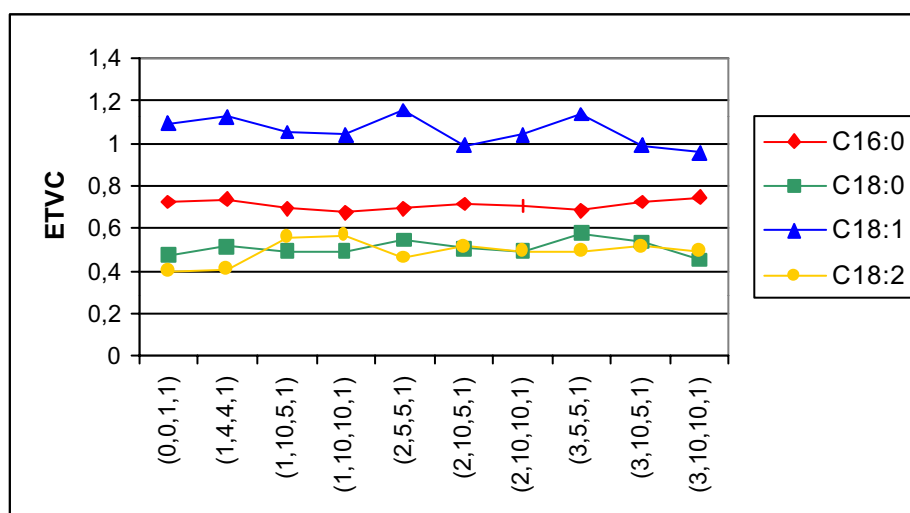
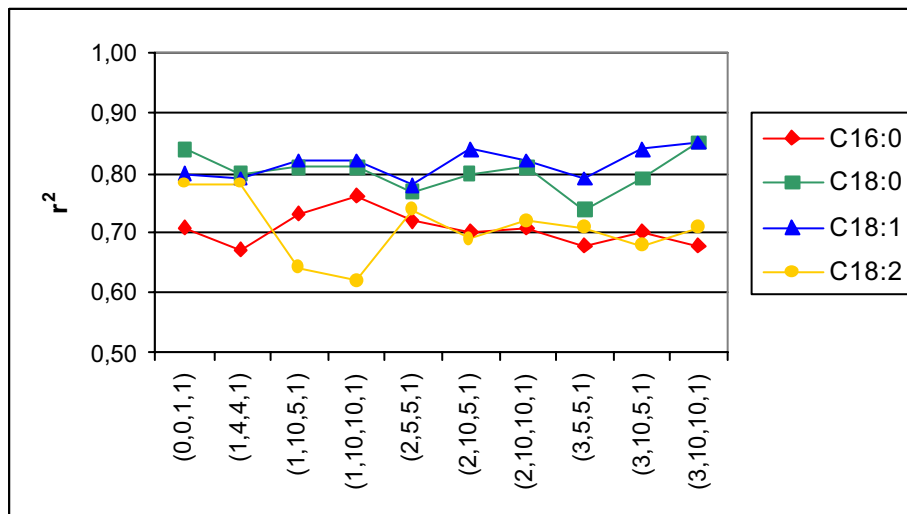


Figura 4.3.41. Valores de r^2 obtenidos para ecuaciones desarrolladas con diferente tratamiento de derivadas.



Así, los valores de *ETVC* fueron inferiores a 1,20% en todos los ácidos grasos, oscilando entre 0,40 y 0,75% para los ácidos grasos C16:0, C18:0 y C18:2 y entre 0,96 y 1,16% para el %C18:1. Los valores de r^2 variaron entre 0,74 y 0,85 para los ácidos C18:0 y C18:1 y entre 0,62 y 0,78 para los ácidos C16:0 y C18:2.

Estos resultados se confirman al calcular los estadísticos *RPD* y *RER*., indicados en la Tabla 4.3.33.

Tabla 4.3.33. Estadísticos *RPD* y *RER* de las ecuaciones desarrolladas a partir de diferentes tratamientos de derivadas.

Ecuación	Tratamiento derivada	RPD				RER			
		C16:0	C18:0	C18:1	C18:2	C16:0	C18:0	C18:1	C18:2
1	0,0,1,1	1,9	2,4	2,3	2,2	11,2	14,7	13,6	13,0
2	1,4,4,1	1,8	2,2	2,2	2,1	10,6	13,2	13,2	12,9
3	1,10,5,1	1,9	2,3	2,4	1,7	11,6	13,8	14,3	10,0
4	1,10,10,1	2,0	2,3	2,4	1,6	12,1	13,8	14,3	9,8
5	2,5,5,1	1,9	2,1	2,1	2,0	11,3	12,6	12,8	11,9
6	2,10,5,1	1,8	2,2	2,5	1,8	11,1	13,4	14,8	10,7
7	2,10,10,1	1,9	2,3	2,3	1,9	11,2	13,6	14,0	11,4
8	3,5,5,1	1,8	1,9	2,2	1,9	10,5	11,7	13,0	11,2
9	3,10,5,1	1,8	2,1	2,5	1,8	10,9	12,9	14,9	10,7
10	3,10,10,1	1,8	2,6	2,6	1,9	10,6	15,5	15,6	11,2

Como se aprecia en dicha tabla, todos los valores del estadístico *RER* fueron superiores al valor mínimo recomendado (Williams y Sobering, 1993), a excepción de la ecuación desarrollada para el ácido C18:2 con un tratamiento de derivada (1,10,10,1) que fue ligeramente inferior (*RER*=9,8). En el caso del estadístico *RPD*, los valores

obtenidos se aproximaron al mínimo recomendado, pero no se llegó a alcanzar para ninguna de las ecuaciones de ácidos grasos desarrolladas. Así, los valores de *RPD* para los ácidos grasos C18:0 y C18:1 oscilaron entre 1,9 y 2,6 mientras que para C16:0 y C18:2 variaron entre 1,6 y 2,2 lo cual indicaba una ligera menor precisión de las ecuaciones obtenidas para estos últimos ácidos grasos.

De acuerdo con los resultados obtenidos en la Tabla 4.3.33, sigue existiendo una discrepancia en cuanto a la superación de los valores mínimos recomendados por los estadísticos *RER* y *RPD*, independientemente del tipo de tratamiento de derivadas empleado. De nuevo, este hecho podría explicarse por las características de las curvas de distribución de los histogramas de composición de ácidos grasos ya comentadas con anterioridad.

Se ha de precisar que el número de muestras finales del colectivo de calibración consideradas para obtener las diferentes ecuaciones NIRS osciló entre un mínimo de 146 y un máximo de 161, dependiendo de la variable y derivada considerada. Sin embargo, en ningún caso el número de muestras eliminadas como anómalas por sus valores de los estadísticos *H* y *T* supuso más del 20% del número inicial de muestras a emplear en calibración, límite recomendado (Shenk y Westerhaus, 1995b) en el desarrollo de ecuaciones de calibración NIR.

Para evaluar el efecto del diferente tratamiento de derivadas sobre la precisión de las ecuaciones, se empleó de nuevo el criterio expuesto en el apartado 4.3.1.3 para la comparación de ecuaciones de calibración NIRS, basado en el protocolo de selección de ecuaciones descrito por Windham *et al.* (1989). Se ha de recordar que según este criterio, diferentes ecuaciones de calibración NIRS se considerarán con la misma precisión cuando los valores de *ETVC* no difieran entre ellas en más de un 20%. Para ello, se han calculado, para cada ácido graso, la máxima diferencia entre los valores de *ETVC* obtenidos de las ecuaciones con diferente tratamiento matemático a partir de la fórmula ya descrita en el apartado 4.3.1.3.

Los valores de los *ETVC* mínimo, máximo y de la diferencia máxima calculada entre los estadísticos *ETVC* de las ecuaciones con diferentes tratamientos de derivada para cada ácido graso aparecen reflejados en la Tabla 4.3.34.

Tabla 4.3.34. Valores de ETVC máximos, mínimos y diferencias máximas en este estadístico para las ecuaciones de cada ácido graso.

Ac. graso	ETVC mínimo	ETVC máximo	Diferencia máxima ETVC (%)
C16:0 (%)	0,67	0,75	12
C18:0 (%)	0,45	0,58	29
C18:1 (%)	0,96	1,16	21
C18:2 (%)	0,40	0,57	43

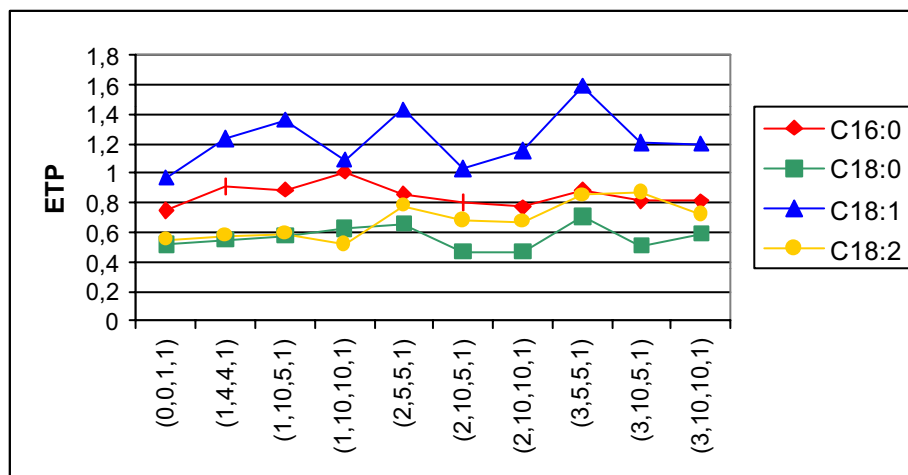
Como puede observarse en dicha tabla, la diferencia máxima en el estadístico *ETVC* para el ácido graso C16:0 (12%) fue inferior al valor máximo recomendado. Sin embargo, estas diferencias superaron el valor máximo para los ácidos grasos C18:0 (29%) y C18:1 (21%) y en particular para el ácido C18:2 donde esta diferencia máxima fue del 43%. Por ello, teniendo en cuenta los valores de *ETVC*, parece existir ligeras diferencias entre los tratamientos de derivadas empleados, si bien estas no son suficientemente claras para todos los ácidos grasos. Sería necesario profundizar en estas diferencias mediante la aplicación de otras metodologías recogidas en la bibliografía (Fearn, 1996; Roggo *et al.*, 2002; Næs *et al.*, 2002).

A continuación se procedió a corroborar su precisión mediante la validación externa de las ecuaciones desarrolladas mediante RLM con el colectivo de validación D. La Figura 4.3.42 representa los valores de *ETP* de los cuatro ácidos grasos para el colectivo de validación D, tras emplear las ecuaciones NIRS obtenidas con los distintos tratamientos de derivadas considerados.

En dicha figura se puede observar que los valores de *ETP* para cada uno de los ácidos grasos considerados presentaron unos intervalos de variación ligeramente más amplios que los respectivos valores de *ETVC* (Figura 4.3.40). Así, los valores de *ETP* oscilaron entre valores similares a los valores de *ETVC* hasta valores un 30% superiores a los valores de *ETVC* obtenidos previamente (entre 0,75 y 1,01% para C16:0; entre 0,47 y 0,71% para C18:0; entre 0,97 y 1,59% para C18:1; entre 0,52 y 0,87% para C18:2).

Además, se ha de indicar que los valores de *ETP* en las predicciones de un mismo ácido graso afectaron de forma similar a los valores de *sesgo* y *ETP(C)*, superándose en alguna ocasión los valores límite recomendados por las distintas ecuaciones para dichos estadísticos.

Figura 4.3.42. Valores de ETP del colectivo D obtenidos con ecuaciones desarrolladas con diferente tratamiento de derivadas.



Por tanto y a modo de conclusión indicar que tras la evaluación externa de las ecuaciones desarrolladas con diferente tratamiento de derivadas, se detectaron de nuevo diferencias entre la precisión de las ecuaciones obtenidas, si bien estas no fueron tan claras como las observadas en las ecuaciones desarrolladas en grasa fundida.

4.3.2.4 Obtención de ecuaciones para la determinación de ácidos grasos minoritarios

Al igual que se expuso en el apartado 4.3.1.8, se desarrollaron ecuaciones de calibraciones NIRS para aquellos ácidos grasos presentes en bajas concentraciones (minoritarios) en tejido adiposo de cerdo Ibérico (C14:0, C17:0, C17:1, C18:3, C20:0 y C20:1).

Para el desarrollo de estas ecuaciones de calibración se han empleado los datos de referencia obtenidos por cromatografía de gases y los datos espectroscópicos correspondientes a muestras del colectivo de calibración descrito en el apartado 4.3.2.1. La Tabla 4.3.35 muestra la composición media, los valores máximo y mínimo y la desviación típica para cada uno de los ácidos grasos minoritarios del colectivo de calibración.

Los datos de dicha tabla muestran que, al igual que ocurría en el caso de grasa fundida, cada uno de los ácidos grasos minoritarios se encuentran en el tejido adiposo a una concentración inferior al 3%, oscilando entre una composición media del

2,20% para el ácido C16:1 y del 0,22% para el ácido C20:0. Además, los intervalos de variación y *DT* son muy inferiores a los existentes en ácidos grasos mayoritarios (Tabla 4.3.27).

Tabla 4.3.35. Composición en ácidos grasos minoritarios de las muestras empleadas como colectivo de calibración (n=164).

Ácido graso	Media	DT	Mínimo	Máximo
C14:0 (%)	1,37	0,14	1,03	1,76
C16:1 (%)	2,20	0,31	1,59	3,20
C17:0 (%)	0,33	0,06	0,20	0,50
C17:1 (%)	0,31	0,08	0,20	0,50
C18:3 (%)	0,63	0,13	0,44	1,10
C20:0 (%)	0,22	0,04	0,15	0,33
C20:1 (%)	1,60	0,18	1,10	2,10

En la Figura 4.3.43 se muestran los diagramas de frecuencias del contenido en ácidos grasos minoritarios del colectivo de calibración.

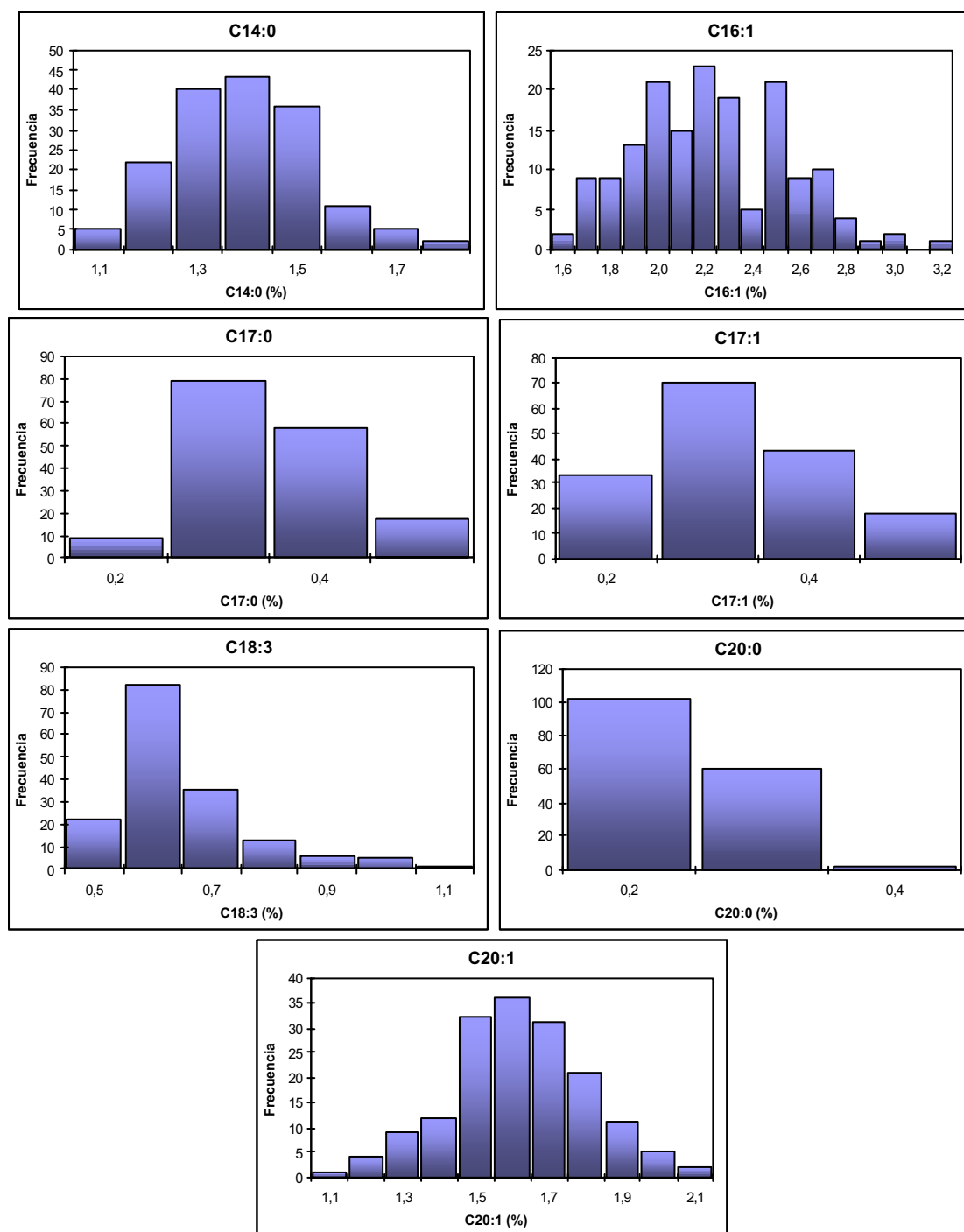
Tras observar los datos incluidos en la Tabla 4.3.35 y la Figura 4.3.43, se observa que, al igual que en el estudio realizado en grasa fundida, las muestras del colectivo de calibración permitieron cubrir la práctica totalidad el intervalo de variabilidad existente en estos ácidos grasos minoritarios para tejido adiposo de cerdo Ibérico.

Al igual que ocurría en grasa fundida, los ácidos grasos presentes en menor concentración (C17:0, C17:1 y C20:0) presentan valores de composición prácticamente constantes, siendo muy pequeños tanto sus intervalos de variación como sus valores de desviaciones típicas (inferiores a 0,08).

Con los datos espectrales y químicos de las muestras de este colectivo se procedió a la obtención de diferentes ecuaciones de calibración para cada uno de los siete ácidos grasos minoritarios considerados, haciendo uso de nuevo de la secuencia automática AUTOCAL, descrita en el apartado 3.3.1.4.

De entre todas las ecuaciones generadas se seleccionaron aquellas con mayor exactitud y precisión de acuerdo con los estadísticos r^2 y *ETVC*.

Figura 4.3.43. Diagramas de distribución de frecuencias del contenido en ácidos grasos minoritarios del colectivo de calibración.



La Tabla 4.3.36 muestra los estadísticos obtenidos para las ecuaciones seleccionadas. Estas ecuaciones se obtuvieron a partir de un tratamiento de derivadas de (1,4,4,1) para el ácido C14:0, (1,10,10,1) para los ácidos C17:0 y C17:1, (2,5,5,1) para C20:0, (2,10,5,1) para C16:1, (2,10,10,1) para C18:3 y (3,5,5,1) para el ácido graso C20:1. El número de términos RMCP seleccionados mediante validación cruzada para

cada ecuación fue de 1 para C20:1, 2 para C20:0, 3 para C14:0, 4 para C16:1 y C17:0, 6 para C17:1 y 8 para C18:3.

Tabla 4.3.36. Estadísticos de ecuaciones NIRS para ácidos grasos minoritarios obtenidas a partir del colectivo de calibración (n=164).

Ac. graso	Media	DT	ETL ⁽¹⁾	ETC	R ²	ETVC	r ²	RPD	RER
C14:0 (%)	1,37	0,14	0,07	0,11	0,36	0,11	0,32	1,20	7,20
C16:1 (%)	2,19	0,30	0,10	0,23	0,38	0,26	0,24	1,15	6,87
C17:0 (%)	0,32	0,05	0,06	0,05	0,20	0,05	0,07	1,03	6,19
C17:1 (%)	0,30	0,07	0,03	0,06	0,38	0,06	0,30	1,19	7,11
C18:3 (%)	0,60	0,09	0,03	0,06	0,61	0,07	0,43	1,32	7,93
C20:0 (%)	0,22	0,04	0,01	0,03	0,36	0,03	0,30	1,19	7,13
C20:1 (%)	1,59	0,17	0,07	0,16	0,16	0,17	0,04	1,02	6,13

1: calculado a partir del análisis por duplicado de 20 muestras incluidas en la calibración

Como puede apreciarse en dicha tabla, los valores de los coeficientes de determinación tanto para la etapa de calibración (R^2) como para la de validación cruzada (r^2) fueron bajos. En ninguna de las ecuaciones, los valores de r^2 fueron superiores a 0,50, límite indicado por Shenk y Westerhaus (1996) para considerar a una ecuación como de alta precisión. Además, los valores obtenidos de *ETVC* para las distintas ecuaciones generadas fueron superiores a los errores de laboratorio (*ETL*) del método de referencia, a excepción de los ácidos grasos C14:0 y C17:0. La baja precisión de las ecuaciones desarrolladas y la similitud entre los valores de *ETL* y *DT* también afecta a los estadísticos *RER* y *RPD*, no superándose los valores mínimos recomendados para estos estadísticos en ninguno de los casos.

Los bajos valores de r^2 obtenidos, en comparación con las ecuaciones descritas para los ácidos grasos mayoritarios (Tablas 4.3.28), también se podrían explicar si se tiene en cuenta la ecuación descrita por Mark y Workman (1991) y expuesta en el apartado 4.3.1.8.

Así, los valores máximo de R^2 que se podrían obtener a partir del colectivo de calibración serían de 0,75 para C14:0, 0,89 para C16:1, 0 para C17:0, 0,82 para C17:1, 0,89 para C18:3, 0,94 para C20:0 y 0,83 para 20:1. Al comparar con los obtenidos para el estadístico R^2 en la Tabla 4.3.36, se puede apreciar los valores descritos con anterioridad son claramente inferiores al límite teórico en las ecuaciones descritas para todos los ácidos grasos.

Por ello, la estrategia a seguir para incrementar la exactitud y precisión de las ecuaciones desarrolladas en este estudio podría ser la minimización del error de laboratorio (*ETL*) y, sobre todo, el error espectral dada la baja repetibilidad de los datos espectroscópicos al emplear la sonda de fibra óptica para analizar tejido adiposo subcutáneo. Además, se recomienda el incrementar el número de muestras en el colectivo de calibración e incrementar, en la medida de los posible, su intervalo de composición.

4.4 DESARROLLO Y EVALUACIÓN DE MODELOS MULTIVARIANTES DE CLASIFICACIÓN DE GRASA DE CERDO IBÉRICO

En los anteriores apartados se han realizado diferentes estudios orientados a corroborar el potencial de la tecnología NIRS como técnica para predecir el contenido en ácidos grasos en grasa de cerdo Ibérico, tanto fundida como subcutánea, para poder estimar el régimen alimenticio durante la etapa de cebo.

En este apartado, el objetivo que se propone es bien diferente. Consiste en mostrar la utilidad de esta técnica como herramienta de clasificación empleando para ello únicamente la información espectral NIR sin necesidad de recurrir a predicciones de analitos como son los ácidos grasos.

Para ello se generarán modelos quimiométricos tanto en grasa fundida como en tejido adiposo mediante diferentes métodos de análisis cualitativo como el análisis discriminante, SIMCA o discriminante basado en regresión RMCP. Se utilizarán como variables tanto el contenido en ácidos grasos, obtenidos mediante predicciones NIRS, como únicamente los datos espectrales NIR tanto en grasa fundida como en tejido adiposo.

Finalmente se compararán los resultados obtenidos en los modelos quimiométricos desarrollados a partir de la composición en ácidos grasos con aquellos de los modelos generados a partir de la información NIR *per se*.

4.4.1 PREDICCIÓN DEL CONTENIDO EN ÁCIDOS GRASOS

Como se ha indicado en el apartado 3.4.1, para el desarrollo de modelos cualitativos se empleó un colectivo formado por muestras individuales de grasa fundida y tejido subcutáneo de 5 lotes de animales con diferente régimen alimenticio durante su etapa de cebo.

Se obtuvo un espectro NIR, tanto de grasa fundida como de tejido adiposo subcutáneo, a partir de cada una de las muestras de este colectivo de aprendizaje, haciendo uso de las metodologías expuestas en los apartados 4.2.1 y 4.2.2 respectivamente. Además, de cada uno de los lotes se obtuvo una muestra de grasa representativa de cada lote, a la cual se le determinó el contenido en ácidos grasos mediante cromatografía de gases.

Sin embargo, no se dispuso de la composición individual de cada una de las muestras individuales que componían los lotes experimentales con diferentes alimentaciones. Dado que se tenía como objetivo el poder comparar los resultados de clasificación obtenidos mediante modelos multivariantes desarrollados a partir de la composición en ácidos grasos o datos espectrales NIR *per se*, para un mismo colectivo, se obtuvieron las predicciones de ácidos grasos (mayoritarios y minoritarios) del colectivo de aprendizaje a partir de estos datos espectrales NIR (en grasa fundida y tejido adiposo) y haciendo uso de las ecuaciones para la determinación del contenido en ácidos grasos descritas en los apartados 4.3.1.6, 4.3.1.8, 4.3.2.1 y 4.3.2.4.

En la Tabla 4.4.1 se incluyen los valores medios de ácidos grasos para cada partida, obtenidos a partir de la predicción individual de los espectros NIR de grasa fundida y tejido adiposo subcutáneo de cada muestra. En dicha tabla también se indican los valores de composición de ácidos grasos, obtenidos mediante cromatografía de gases, de cada una de las partidas. Esta composición corresponde a la de una muestra de grasa fundida media representativa de cada partida, obtenida de acuerdo a la metodología descrita en el apartado 3.1.1.

Como puede apreciarse en la Tabla 4.4.1, las partidas empleadas como colectivo de aprendizaje cubren un amplio intervalo de variabilidad, en cuanto a composición de ácidos grasos se refiere. En concreto, para los ácidos grasos

mayoritarios (C16:0, C18:0, C18:1 y C18:2), parámetros que se incluyen en el contrato tipo homologado (CTH) para determinar la categoría comercial de partidas de cerdos Ibéricos, se observa que todos ellos cubren el intervalo de composición habitualmente descrito por la bibliografía (De Pedro, 1995 y 2000; Benito *et al.*, 1998; Ruiz *et al.*, 1998).

Así, particularizando en el ácido C18:1, la composición oscila entre un valor de 57,3% para el lote 1 (determinado por CG) y un valor de 47,6% para el lote 5. Estos resultados de composición eran previsibles teniendo en cuenta la información de campo disponible acerca del régimen alimenticio empleado en cada lote (De Pedro, 1995 y 2000).

Al comparar los valores medios de composición en ácidos grasos para cada lote, obtenidos mediante los 3 métodos empleados (cromatografía de gases, análisis NIRS de grasa fundida y de tejido adiposo), se aprecia la similitud entre los resultados. Es de mención resaltar los resultados obtenidos para los ácidos grasos mayoritarios. Así, las máximas diferencias entre un valor medio predicho por NIRS (en grasa fundida o tejido adiposo) y el valor de referencia de cromatografía de gases fueron de 0,7 para el ácido C16:0 (CG vs. NIRS en grasa fundida para el lote 4), 1,0 para C18:0 (CG vs. NIRS en grasa fundida para el lote 1), 1,7 para C18:1 (CG vs. NIRS en tejido adiposo para el lote 2) y 0,70 para C18:2 (CG vs. NIRS en tejido adiposo para el lote 5).

Como ya se indicó en el apartado 4.1.3, las máximas diferencias absolutas (para un 95% de los casos) entre 2 resultados de composición, y sólo 2, obtenidos sobre una misma muestra, con la misma metodología analítica, en 2 laboratorios diferentes viene definido por la reproducibilidad. De acuerdo con los resultados expuestos en el apartado 4.1.3, los valores de reproducibilidad para los ácidos grasos mayoritarios fueron de 2,00 para C16:0, 1,05 para C18:0, 2,47 para C18:1 y 0,75 para C18:2.

Tabla 4.4.1. Composición en ácidos grasos (%) de las muestras empleadas en el desarrollo de los modelos.

LOTE	Análisis	C14:0	C16:0	C16:1	C17:0	C17:1	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1	CTH
1 (n = 41)	NIRS Gr. Fun. ¹	1,3	19,0	2,2	0,3	0,4	7,4	57,3	9,8	1,1	0,1	1,5	B
	NIRS T. Adip. ²	1,2	18,8	2,3	0,4	0,4	8,3	57,0	9,6	0,6	0,2	1,9	B
	CG ³	1,2	18,9	2,0	0,2	0,3	8,4	57,3	9,2	1,0	0,1	1,5	B
2 (n = 36)	NIRS Gr. Fun. ¹	1,3	20,4	2,0	0,3	0,4	8,4	54,6	9,7	1,2	0,1	1,3	B
	NIRS T. Adip. ²	1,3	20,4	2,2	0,4	0,4	9,3	53,3	10,2	0,6	0,2	1,7	R
	CG ³	1,3	20,5	1,9	0,2	0,2	8,8	55,0	9,6	1,0	0,1	1,3	B
3 (n = 20)	NIRS Gr. Fun. ¹	1,3	21,3	1,9	0,3	0,4	10,5	53,1	8,5	1,0	0,1	1,6	C
	NIRS T. Adip. ²	1,4	21,7	2,3	0,3	0,4	10,3	52,2	8,9	0,5	0,2	1,6	R
	CG ³	1,2	21,2	1,6	0,2	0,2	11,1	53,4	8,4	0,8	0,2	1,7	C
4 (n = 11)	NIRS Gr. Fun. ¹	1,3	20,2	2,2	0,3	0,4	8,8	53,8	10,7	1,1	0,2	1,5	C
	NIRS T. Adip. ²	1,2	19,9	2,2	0,4	0,4	9,5	53,0	10,9	0,7	0,2	1,8	C
	CG ³	1,3	19,5	1,9	0,3	0,3	9,1	53,7	11,1	1,2	0,1	1,4	C
5 (n = 13)	NIRS Gr. Fun. ¹	1,4	24,1	2,2	0,3	0,4	12,2	47,8	9,2	0,9	0,2	1,4	C
	NIRS T. Adip. ²	1,5	23,6	2,2	0,4	0,4	12,3	47,5	9,5	0,6	0,2	1,5	C
	CG ³	1,4	24,0	2,3	0,4	0,4	12,8	47,6	8,8	0,8	0,2	1,3	C

1: Valores medios de las predicciones NIRS de las muestras individuales de cada lote a partir de espectros de grasa fundida.

2: Valores medios de las predicciones NIRS de las muestras individuales de cada lote a partir de espectros de tejido adiposo.

3: Composición de la muestra media representativa de cada lote obtenido por cromatografía de gases.

En el caso de las máximas diferencias descritas en la Tabla 4.4.1, se han realizado los análisis en 2 laboratorios, empleado muestras diferentes e incluso distintas técnicas analíticas. Por tanto, las máximas diferencias entre los valores de referencia (CG) y los valores predichos NIRS deberían de ser, a priori, superiores a los valores de reproducibilidad, dado que se introducen más fuentes de error (distintas muestras y técnicas analíticas).

A pesar de ello, las máximas diferencias detectadas entre los valores de referencia y los valores medios predichos por NIRS, para los ácidos grasos mayoritarios, fueron inferiores a los valores de reproducibilidad. Es decir, estas diferencias fueron del mismo orden o incluso inferiores que las existentes entre el análisis, mediante cromatografía de gases, de una misma muestra en 2 laboratorios diferentes.

Además, los valores de composición en los ácidos grasos minoritarios, obtenidos mediante cromatografía o estimados a partir del valor medio de los predichos por NIRS en grasa fundida o tejido adiposo fueron similares. Únicamente se apreciaron diferencias elevadas en el valor de C16:1 para el lote 3 (diferencia máxima absoluta entre CG y predicho NIRS en tejido adiposo de 0,7%) y en los valores de C18:3, para todos los lotes, entre los valores de referencia y los estimados mediante los valores medios predichos a partir del espectro NIRS de tejido adiposo (diferencia máxima absoluta de 0,5% para el lote 4).

De acuerdo con los resultados obtenidos en la Tabla 4.4.1 y teniendo en cuenta los estadísticos de las ecuaciones descritas en los apartados 4.3.1.6, 4.3.1.8, 4.3.2.1 y 4.3.2.4 se puede concluir que los valores de composición de ácidos grasos predichos para cada una de las muestras de los distintos lotes, a partir de sus espectros NIRS de grasa fundida o tejido adiposo, no deben diferir de su composición en ácidos grasos, determinada mediante cromatografía de gases.

De forma complementaria a la obtención de la composición en ácidos grasos, se propuso el clasificar la composición media de cada lote independientemente de la técnica analítica empleada. Para ello se aplicaron los criterios analíticos establecidos en el último contrato tipo homologado publicado en el Boletín Oficial del

Estado (BOE, 2000), esto es, los niveles de composición de los ácidos grasos mayoritarios. Dicha clasificación se incluye en la Tabla 4.4.1.

Así, considerando la composición de la muestra media representativa analizada mediante cromatografía de gases, el lote 1 y 2 resultarían clasificados dentro de la categoría comercial *Bellota*, mientras que los lotes 3, 4 y 5 en la categoría de *Cebo*. Si se compara esta clasificación con la información de campo disponible, puede apreciarse el error de clasificación cometido en los lotes 3 y 4.

Al clasificar empleando los valores predichos NIR de ácidos grasos, el lote 1 resultaría clasificado dentro de la categoría *Bellota* y los lotes 4 y 5 en la de *Cebo*. En el caso del lote 4, su clasificación en la categoría comercial de *Cebo* sería como consecuencia del valor elevado del ácido C18:2 (superior a 10,5) mientras que para el lote 5 su clasificación en *Cebo* es debido al incumplimiento de los niveles exigidos para los ácidos C16:0, C18:0 y C18:1.

Los elevados valores de C18:2 del Lote 4 podrían ser debidos al tipo de pienso empleado en la alimentación aportada a los animales como complemento a la montanera. Sin embargo, el largo periodo de aprovechamiento de la montanera le aportó a este lote un perfil de ácidos grasos diferentes al de aquellos animales que sólo se alimentaron a base de pienso durante su engorde. Por esta razón, su calidad sería superior a la resultante de su clasificación según el CTH, siendo demasiado estricta su clasificación en la categoría comercial de *Cebo*.

En el caso del lote 3, según la clasificación de campo (2 meses de montanera con 1 mes con un aporte de 0,5 kg de pienso por día y por animal y el resto sin aporte de pienso) se ha de asignar a una categoría comercial de *Bellota*. Sin embargo, su elevada composición en C18:0 (11,1%) obliga a clasificar el lote, al aplicar los criterios del CTH sobre la composición en ácidos grasos obtenida por cromatografía de gases y predichos en grasa fundida, en la categoría de *Cebo* (composición C18.0 > 10,5%). En el caso de emplear las predicciones de ácidos grasos a partir de espectros de tejido adiposo, el lote resultaría clasificado dentro de la categoría de *Recebo*.

Finalmente, el Lote 2 estaría también próximo a la calidad *Recebo* dado que tuvo un aporte de 0,5 kg/cerdo diarios de pienso durante 1 mes y ningún aporte de pienso durante los 3 meses restantes hasta su sacrificio. Según la aplicación del CTH al

perfil de ácidos grasos obtenidos mediante el análisis NIRS sobre tejido adiposo, dicho lote resultaría clasificado dentro de la calidad *Recebo* mientras que mediante el análisis de grasa fundida por cromatografía de gases y predicciones NIRS en grasa fundida sería asignado a la calidad *Bellota*.

Por tanto, se observa que existen errores de importancia entre la clasificación de campo y la realizada al aplicar los criterios del CTH sobre la composición de ácidos grasos determinados mediante cromatografía de gases. Además, incluso en algunas partidas existen discrepancias entre la clasificación realizada en función de la técnica analítica empleada para determinar el contenido en ácidos grasos.

En este último caso, las discrepancias en cuanto a clasificación vendrían provocadas no por las diferencias en cuanto a composición obtenida mediante las diferentes técnicas, sino por los propios criterios analíticos de clasificación del CTH. Ello se debería a que, teniendo en cuenta los resultados descritos en el apartado 4.1.3, las diferencias obtenidas en composición de ácidos grasos por las distintas técnicas son del mismo orden e incluso inferiores a los valores de reproducibilidad del análisis de ácidos grasos mediante cromatografía de gases. Es decir, estas discrepancias en cuanto a valores de composición y, por tanto, en su posterior clasificación, también se podrían haber observado en el caso de que únicamente se hubiera empleado la cromatografía de gases como técnica de análisis en dos laboratorios diferentes.

Ante esta situación, en los siguientes apartados se van a estudiar otras posibles estrategias de clasificación mediante técnicas multivariantes para evaluar el régimen alimenticio a partir tanto de la composición en ácidos grasos predichos mediante NIRS o bien los datos espectrales NIRS *per se* de las muestras de grasa fundida y tejido adiposo de los lotes anteriormente descritos.

4.4.2 DESARROLLO DE MODELOS CUALITATIVOS MEDIANTE ANÁLISIS DISCRIMINANTE

Como ya fue expuesto en el apartado 3.4.2.1, el objetivo del análisis discriminante es comprobar en qué medida es posible diferenciar correctamente las clases establecidas en el grupo de aprendizaje a partir de las variables analizadas, estableciéndose criterios estadísticos que estimen la distancia entre las clases y asignen

objetos a la clase correspondiente (Mark, 1992; Massart *et al.*, 1988 y 1997; Blanco *et al.*, 1998; Downey, 1996 y 2000). Para una muestra desconocida, se procederá al cálculo de la distancia de cada muestra al centroide de las diferentes clases, siendo clasificada en aquel grupo al cual presente menor distancia o mayor probabilidad de pertenencia a una clase (Mark, 1992; Massart *et al.*, 1998 y 1997).

Como se ha indicado en el apartado 3.4.1.3, para cada método de clasificación multivariante a evaluar, se desarrollarán modelos utilizando como variables de entrada la composición en ácidos grasos predicha por NIRS para grasa fundida y tejido adiposo y modelos empleando como variables la información espectral NIR *per se* de las mismas muestras.

El análisis discriminante puede ser empleado sobre las variables consideradas o sobre las mismas sintetizadas mediante algún método multivariante como el análisis de componentes principales.

Dada la colinealidad existente entre en los valores de absorbancia de la región NIR y asimismo la alta correlación entre los datos de composición de ácidos grasos, se realizó un análisis de componentes principales sobre cada una de estas variables como tratamiento previo de la información. Dicho análisis de componentes principales se realizó tanto en los modelos cualitativos generados a partir de muestras de tejido adiposo como de grasa fundida.

De acuerdo con lo descrito en el apartado 3.4.2.1, se han evaluado 3 métodos diferentes de análisis discriminante tales como el análisis discriminante canónico, análisis discriminante lineal y análisis discriminante cuadrático.

4.4.2.1 *Análisis discriminante canónico*

Como ya se indicó en el apartado 3.4.2.1, mediante el análisis discriminante canónico se calculan unas nuevas variables (variables canónicas discriminantes) que son combinación lineal de las variables originales (componentes principales de los datos espectrales o de los valores predichos de ácidos grasos), elegidas de tal manera que se minimice la variabilidad de las muestras de un mismo lote y se maximice la variabilidad entre lotes.

El criterio empleado para elegir el número de componentes principales fue el que explicaran un mínimo del 99,9% de la varianza inicial. Para alcanzar dicho valor de la varianza explicada fueron necesarias 5 componentes principales, tanto para el modelo desarrollado con ácidos grasos en grasa fundida como en tejido adiposo, y 9 componentes principales para el modelo generado a partir de datos espectrales NIR *per se* en grasa fundida.

En el caso del modelo que empleaba como variables los datos NIRS en tejido adiposo, ni siquiera el número máximo de componentes principales calculadas (máximo = 15) permitió explicar el 99,9% de la varianza inicial. Por ello, en este caso se seleccionaron 10 componentes principales para el desarrollo del modelo, siendo la varianza explicada de un 98,3%.

En el caso de los datos de composición en ácidos grasos, el hecho de que se seleccionara un mismo número de componentes principales que explicaran el modelo indicaría que las variables originales utilizadas (ácidos grasos predichos mediante NIRS) producirían una definición similar (variabilidad similar) del colectivo de aprendizaje a emplear en el análisis discriminante.

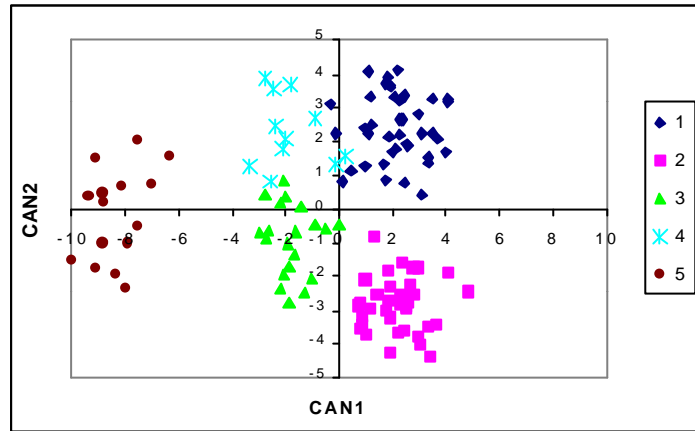
El mayor número de componentes principales necesarias para explicar la misma variabilidad del colectivo inicial en base a la información espectroscópica NIR, podría indicar que las variables definidas a partir de los valores de absorbancia *per se*, contienen una mayor información sobre la variabilidad de dicho colectivo que la información obtenida a partir de la composición en ácidos grasos.

La necesidad de un mayor número de componentes principales para explicar la variabilidad del colectivo de aprendizaje a partir de datos NIRS en tejido adiposo que en grasa fundida indicaría una mayor variabilidad del colectivo obtenido en tejido adiposo. Esto podría ser en parte explicado por el mayor error de la medida espectral, bien por una menor repetibilidad (apartado 4.2.1 y 4.2.2) o bien por la introducción de alguna fuente de error adicional mediante el uso de la sonda de fibra óptica.

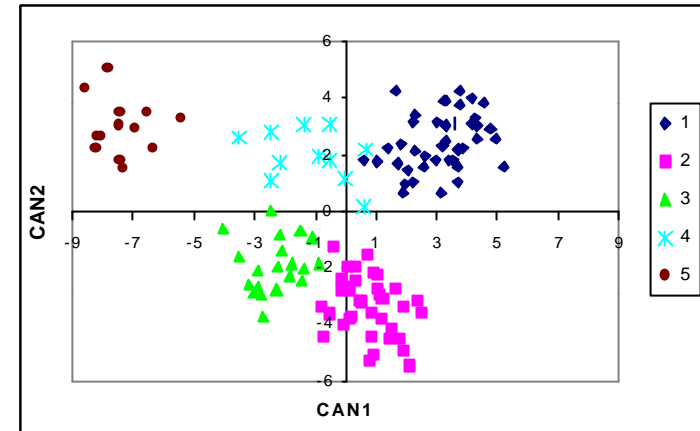
En la Figura 4.4.1 se representan las muestras del colectivo de aprendizaje en los planos definidos por las dos primeras variables canónicas (CAN1 vs. CAN2) para los cuatro modelos generados mediante análisis discriminante canónico.

Figura 4.4.1. Representación de las muestras en los modelos generados mediante análisis discriminante canónico.

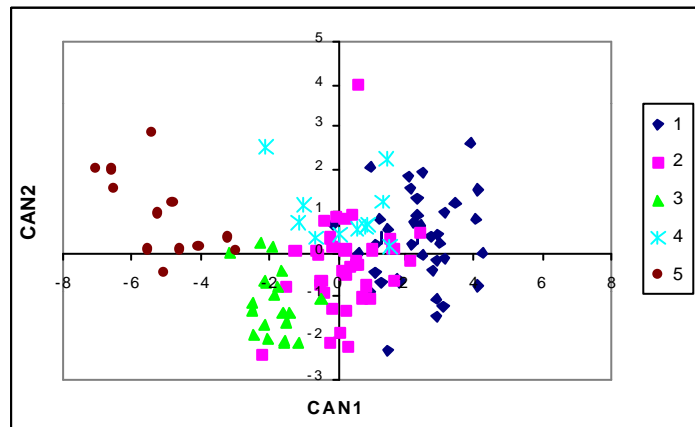
Ácidos grasos de grasa fundida



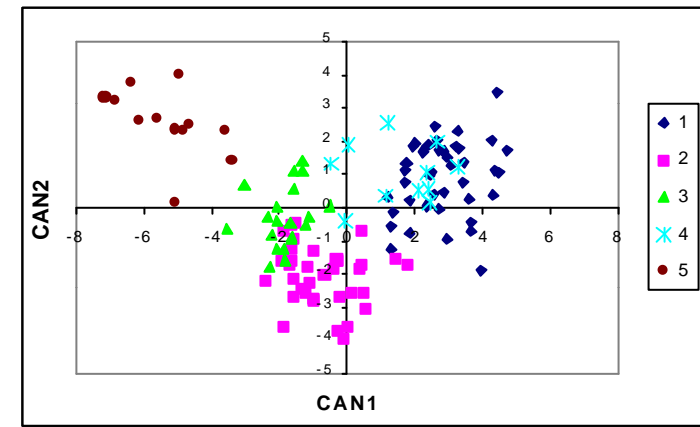
Datos espectrales NIRS per se de grasa fundida



Ácidos grasos de tejido adiposo



Datos espectrales NIRS per se de tejido adiposo



Como puede apreciarse en dicha figura, en todos los modelos generados se aprecia una clara correlación entre el tipo de alimentación empleado durante el cebo y el valor de la primera variable canónica (CAN1). Así, los lotes experimentales con un mayor aporte de bellota durante su cebo (lotes 1 y 2) poseen valores de la variable CAN1 superiores a 0 mientras que los lotes con menor reposición en montanera a base de bellotas (lotes 3, 4 y 5) presentan, en todos los modelos, valores próximos o inferiores a cero. El lote 5, con una alimentación a base de pienso exclusivamente durante la fase de cebo presenta, en todos los modelos, valores de la primera variable discriminadora (CAN1), claramente negativos (inferiores a -3).

Sin embargo, el nivel de diferenciación entre los lotes del colectivo de aprendizaje fue diferente dependiendo del modelo considerado. Así, la mejor discriminación se logró en el modelo obtenido a partir de datos NIRS en grasa fundida donde la separación entre los lotes con diferente alimentación fue casi perfecta. Únicamente se apreció un ligero solape entre los lotes 1 y 4 y entre los lotes 2 y 3, teniendo en cuenta únicamente la información facilitada por las dos primeras variables canónicas.

Este hecho puede ser considerado normal en este tipo de muestras dada la similitud entre la alimentación de los lotes experimentales considerados y la elevada variabilidad animal existente en lotes sometidos a alimentación en extensivo.

Así, el lote 2 y 3 se diferenciaron básicamente en la duración de la montanera, la cual fue 2 meses menor en el lote 3. Puede resultar llamativo el nivel de proximidad existente entre el lote 1 y algunas muestras del lote 4. Sin embargo, teniendo en cuenta la información de campo, estos lotes estuvieron durante un periodo largo en montanera (superior a 6 meses), diferenciándose básicamente en el aporte de pienso por animal.

Así, el lote 1 no tuvo prácticamente ningún aporte de pienso mientras que el lote 4 necesitó un complemento importante de pienso en su alimentación hasta alcanzar el peso de sacrificio (15 @ aprox.). Este aporte de pienso pudo provocar una mayor variabilidad animal en dicho lote, dentro del cual existirían cerdos que podrían haber ingerido únicamente bellota en su dieta diaria mientras que otros podrían haber ingerido diariamente y de forma exclusiva el pienso aportado de forma complementaria. Por ello,

se podría explicar que ciertas muestras de este lote se encuentren próximas al lote 1 mientras que otras se localicen más cerca del lote 5 (alimentado a base de pienso durante el cebo).

La separación entre lotes obtenida con el modelo que ha empleado como variables la predicción del contenido en ácidos grasos de la grasa fundida, fue también elevada (Figura 4.4.1). Sin embargo, comparando estos resultados con los obtenidos en el modelo desarrollado a partir de datos NIRS de la grasa fundida, el nivel de discriminación entre los lotes fue inferior. Así, se aprecia un mayor solape entre los lotes 1 y 4 y entre los lotes 3 y 4 que en el modelo generado con información espectral *per se*.

Este mayor solape entre los lotes 3 y 4 puede deberse a la similitud en sus valores medios de composición de ácidos grasos. Así, la composición en ácidos mayoritarios C18:1 y C16:0 es parecida, siendo mayores las diferencias (del 2% aprox.) en los ácidos C18:0 y C18:2. Es por ello que alguna de las muestra de dichos lotes puedan presentar incluso la misma composición en ácidos grasos, con lo cual, obviamente deberían situarse próximos en el plano definido por las dos primeras variables canónicas.

A pesar de esta semejanza en cuanto a composición en ácidos grasos, las coordenadas de los centroides de los lotes 3 y 4 en el plano definido por las dos primeras coordenadas canónicas son claramente diferentes (en particular, para la variable CAN2) por lo que se corroboran los trabajos realizados por De Pedro (2000) acerca de las amplias posibilidades de uso de técnicas multivariantes para la evaluación del régimen alimenticio en cerdo Ibérico.

En el caso de los lotes 1 y 4, existen diferencias claras en cuanto a composición media de ácidos grasos, motivadas por el diferente régimen alimenticio empleado en cada uno de ellos. Así, a modo de ejemplo, la composición media de C18:1 del lote 1 fue de 57,3% mientras que para el lote 4 fue de 53,7%. Sin embargo, como ya se comentó previamente, el lote 4 podría presentar una elevada variabilidad en cuanto a la composición en ácidos grasos de la grasa, dada el complemento elevado de pienso en su alimentación. Este hecho se confirma al comprobar que, por ejemplo, los valores de composición en C18:1 oscilaron en el mismo lote entre un valor mínimo de

51,3% y un valor máximo de 55,7%. En el caso del lote 1, la variabilidad en cuanto a composición en ácidos grasos fue menor si bien, el valor mínimo de C18:1 predicho para una de las muestras fue de 54,6%, inferior al valor máximo de C18:1 presente en el lote 4.

Este es el caso de la muestra 409 perteneciente al Lote 4 (C16:0=19,6%; C18:0=8,6%; C18:1=55,7%; C18:2=9,8%) y la muestra 107 perteneciente al Lote 1 (C16:0=19,9%; C18:0=8,8%; C18:1=55,7%; C18:2=9,3%). Con ello se aprecia que la composición en ácidos grasos de algunas muestras de los lotes 1 y 4 fueron similares y, por tanto, su posición en el espacio definido por las variables canónicas ha de ser muy cercana, provocando un solape entre dichos lotes.

En los modelos generados a partir de información procedente de tejido adiposo, la diferenciación entre lotes fue menor. Así, en el caso de uso de los datos espectrales NIR como variables, se aprecia una superposición entre ciertas muestras del lote 4 y el lote 1, al proyectar estas en un plano definido por las dos primeras variables canónicas (Figura 4.4.1). Además, se aprecia un claro solape entre las muestras de los lotes 2 y 3, hecho que no se apreciaba en el caso de grasa fundida.

Al igual que en los modelos desarrollados para grasa fundida, con los resultados obtenidos con el modelo generado a partir de valores predichos de ácidos grasos en tejido adiposo se logra una menor diferenciación entre lotes que en el caso de emplear datos NIRS como variables. Así, los lotes 2, 3 y 4 presentan un elevado nivel de solape al considerar las dos primeras variables canónicas, quedando únicamente más diferenciados de ellos los lotes 1 y 5. Por tanto, el modelo tendría una mayor dificultad a la hora de poder discriminar entre muestras de alimentación similar (como es el caso de los lotes 2, 3 y 4) si bien lograría diferenciar muestras de lotes con alimentación muy diferente (lote 1 y 5).

En general y teniendo en cuenta los resultados expuestos en la Figura 4.4.1 del análisis discriminante canónico, se puede concluir que es posible lograr una mejor discriminación entre lotes con diferente alimentación a partir de modelos generados con información procedente de grasa fundida frente a aquellos desarrollados con información de muestras de tejido adiposo.

Un aspecto de gran importancia a resaltar en estos resultados exploratorios es que en todos los modelos generados, el lote 5 (alimentación en el cebo a base exclusivamente de pienso) resultó claramente diferenciado del resto de lotes experimentales, los cuales tuvieron una mayor o menor presencia de la bellota en su régimen alimenticio durante el cebo.

Como ya se indicó en el apartado 3.4.2.1, mediante un análisis discriminante canónico se evalúan las posibilidades de discriminación entre distintas clases, no siendo posible una asignación de muestras a una determinada clase, opción que metodológicamente es abarcable mediante el análisis discriminante lineal o cuadrático (Massart *et al.*, 1997; Næs *et al.*, 2002). La función discriminante desarrollada, ya sea lineal o cuadrática, permite la diferenciación de las muestras en las clases definidas y la clasificación de muestras no incluidas en el desarrollo del modelo.

4.4.2.2 *Análisis discriminante cuadrático*

La distinción entre análisis discriminante lineal y cuadrático se basa en la forma de obtención de la función discriminante cuando la misma se obtiene a partir de una matriz de covarianzas ponderada, dando lugar a una función lineal, o cuando se calcula a partir de las distintas matrices de covarianzas de las clases, generando una función cuadrática (Mark, 1992; Massart *et al.*, 1988 y 1997; Næs *et al.*, 2002).

El emplear un tipo u otro de matrices de covarianzas, y por tanto de análisis discriminante depende de la igualdad o no de las matrices de covarianza. Por ello, un paso previo al análisis discriminante ha consistido en la realización de un test de homogeneidad entre las matrices de covarianzas de cada una de las clases (Morrison, 1976; SAS, 1990), para establecer el tipo de análisis discriminante más idóneo para el colectivo de aprendizaje empleado.

Los resultados obtenidos de dicho test han mostrado que, independientemente del tipo de variables empleadas para el desarrollo de los modelos (componentes principales de valores predichos de ácidos grasos o de datos espectrales NIR) o de la forma de presentación de muestras elegida (grasa fundida o tejido adiposo subcutáneo), el análisis discriminante recomendado ha sido el cuadrático.

Únicamente se recomendó el desarrollo de un análisis discriminante lineal al generar modelos a partir de una única componente principal de valores de ácidos grasos o de datos NIRS procedentes de grasa fundida. Al emplear más de una componente principal en el desarrollo de estos modelos, se recomendó el desarrollo de modelos discriminantes cuadráticos.

Dado que con una única componente principal el porcentaje de varianza explicada fue siempre inferior al 90%, se descartó esta opción, decidiéndose el desarrollo de modelos discriminantes cuadráticos utilizando un mayor número de componentes principales que explicaran un mínimo del 99% de varianza inicial.

Esta recomendación de un análisis discriminante cuadrático frente a uno lineal puede deberse a la propia estructura del colectivo de aprendizaje. Así, este colectivo no tuvo una estructura homogénea, presentando cada clase (lote) un número diferente de muestras que oscilaron entre un valor de 41 para el lote 1 y un valor de 11 para el lote 4. Este hecho podría ser responsable de la desigualdad de las matrices de covarianzas entre las clases y por tanto, la recomendación de las distintas matrices de cada lote para el cálculo de la función discriminante.

Por ello, en la Tabla 4.4.2 se muestran las matrices de clasificación resultantes de los modelos desarrollados mediante análisis discriminante cuadrático a partir de valores predichos de ácidos grasos y de datos espectrales NIR en grasa fundida y tejido adiposo.

El número de componentes principales óptimas (ya sea de valores predichos de ácidos grasos o de datos NIRS) a emplear en el análisis discriminante cuadrático se estimó mediante validación cruzada completa.

En este proceso, cada una de las muestras del colectivo de aprendizaje es clasificada mediante un modelo obtenido a partir del resto de muestra de dicho colectivo. Con ello, se estima un error de clasificación medio más real que el que se obtendría mediante la reclasificación de las propias muestras del colectivo de aprendizaje. Así, se seleccionó aquel número de componentes principales mediante el cual se obtuvo un valor mínimo en el error de clasificación resultante del proceso de validación cruzada completa.

Tabla 4.4.2. Resultados de los modelos desarrollados mediante análisis discriminante cuadrático, a partir de valores predichos de ácidos grasos y datos espectrales NIR.

Ácidos grasos en grasa fundida						Datos NIRs en grasa fundida					
CLASIFICADO EN LOTE						CLASIFICADO EN LOTE					
1 2 3 4 5						1 2 3 4 5					
LOTE ORIGEN	1	40 97,6%	-	-	1 2,4%	-	39 95,1%	2 4,9%	-	-	-
	2	-	36 100,0%	-	-	-	1 2,8%	35 97,2%	-	-	-
	3	-	-	20 100,0%	-	-	-	-	20 100,0%	-	-
	4	1 9,1%	-	-	10 90,9%	-	1 9,1%	-	-	10 90,9%	-
	5	-	-	-	-	13 100,0%	-	-	-	-	13 100,0%
Error de Clasificación: 1,7%						Error de Clasificación: 3,3%					
Ácidos grasos en tejido adiposo						Datos NIRs en tejido adiposo					
CLASIFICADO EN LOTE						CLASIFICADO EN LOTE					
1 2 3 4 5						1 2 3 4 5					
LOTE ORIGEN	1	38 92,7%	2 4,9%	-	1 2,4%	-	31 75,6%	2 4,9%	-	8 19,5%	-
	2	4 11,1%	29 80,6%	-	3 8,3%	-	2 5,5%	32 88,9%	1 2,8%	1 2,8%	-
	3	-	1 5,0%	16 80,0%	-	3 15,0%	-	4 20,0%	16 80,0%	-	-
	4	-	4 36,4%	-	7 63,6%	-	4 36,4%	-	1 9,0%	6 54,6%	-
	5	-	-	2 15,4%	-	11 84,6%	-	2 15,4%	-	-	11 84,6%
Error de Clasificación: 16,5%						Error de Clasificación: 20,7%					

El número de componentes principales seleccionadas para el desarrollo de modelos en grasa fundida fue de 4 al emplear los valores de ácidos grasos como variables y de 5 al considerar los datos espectrales NIR. En el caso de los modelos generados para muestras de tejido adiposo, se seleccionaron 5 componentes principales al emplear como variables los valores de ácidos grasos y 6 al considerar los datos espectrales NIR.

En la Tabla 4.4.3 se indican las muestras mal clasificadas por los modelos desarrollados junto con el valor de probabilidad (%) de pertenencia de cada muestra a cada uno de los lotes (clases).

Los resultados incluidos en dicha tabla confirman aquellos obtenidos mediante análisis discriminante canónico de forma que los modelos desarrollados en

muestra de grasa fundida permiten diferenciar mejor los distintos lotes del colectivo de calibración que aquellos desarrollados con muestras de tejido adiposo. Así, el error de clasificación estimado mediante validación cruzada completa fue inferior al 4% en el caso de los modelos generados en grasa fundida mientras que los modelos desarrollados en tejido adiposo, presentó un valor superior al 15%.

Tabla 4.4.3. Muestras mal clasificadas por los modelos generados mediante análisis discriminante cuadrático, a partir de valores predichos de ácidos grasos y datos espectrales NIR.

Ácidos grasos en grasa fundida							Datos NIRS en grasa fundida								
Muest.	Lote	Clasif.	PROBABILIDAD CLASIFICACIÓN LOTE (%)					Muest.	Lote	Clasif.	PROBABILIDAD CLASIFICACIÓN LOTE (%)				
			1	2	3	4	5				1	2	3	4	5
132	1	4	15,3	0,0	0,0	84,7	0,0	107	1	2	47,7	52,2	0,1	0,0	0,0
409	4	1	99,4	0,0	0,0	0,6	0,0	138	1	2	0,8	99,3	0,0	0,0	0,0
								227	2	1	66,0	34,0	0,0	0,0	0,0
								401	4	1	82,8	14,6	0,0	2,7	0,0
Error de Clasificación: 1,7%							Error de Clasificación: 3,3%								
Ácidos grasos en tejido adiposo							Datos NIRS en tejido adiposo								
Muest.	Lote	Clasif.	PROBABILIDAD CLASIFICACIÓN LOTE (%)					Muest.	Lote	Clasif.	PROBABILIDAD CLASIFICACIÓN LOTE (%)				
			1	2	3	4	5				1	2	3	4	5
114	1	2	13,5	86,5	0,0	0,0	0,0	123	1	2	14,6	79,0	0,0	6,4	0,0
135	1	2	10,8	89,2	0,0	0,0	0,0	124	1	4	13,5	0,2	0,0	86,4	0,0
138	1	4	10,8	8,0	0,0	81,2	0,0	125	1	4	30,6	17,6	0,0	51,9	0,0
201	2	1	60,9	15,0	0,0	24,1	0,0	130	1	2	7,9	92,1	0,0	0,0	0,0
202	2	4	0,5	2,1	0,0	97,4	0,0	131	1	4	47,3	1,0	0,0	51,7	0,0
203	2	4	4,6	17,9	27,9	49,6	0,0	132	1	4	2,7	0,1	0,0	97,3	0,0
204	2	1	46,5	37,7	0,0	15,9	0,0	133	1	4	30,4	0,1	0,0	69,6	0,0
208	2	4	0,6	2,9	0,0	96,5	0,0	134	1	4	21,7	0,0	0,0	78,3	0,0
215	2	1	60,7	36,1	0,0	3,3	0,0	136	1	4	12,7	0,3	0,0	87,0	0,0
223	2	1	74,8	24,2	0,0	1,0	0,0	150	1	4	39,2	4,6	0,0	56,2	0,0
304	3	5	0,0	33,0	17,0	15,0	35,0	201	2	1	99,9	0,1	0,0	0,0	0,0
312	3	2	26,3	65,2	8,4	0,1	0,0	202	2	4	10,5	11,0	0,0	78,5	0,0
314	3	5	0,1	24,3	19,1	3,9	52,6	203	2	3	0,1	5,8	94,1	0,0	0,0
315	3	5	0,0	0,0	0,6	0,3	99,0	204	2	1	74,2	2,6	0,0	23,1	0,0
407	4	2	40,1	59,9	0,0	0,0	0,0	301	3	2	0,0	54,5	45,2	0,0	0,3
410	4	2	0,0	81,2	0,0	0,0	18,8	306	3	2	0,0	75,7	5,6	17,6	1,1
411	4	2	7,8	66,6	0,0	25,6	0,0	312	3	2	1,1	97,2	1,7	0,0	0,0
412	4	2	11,8	87,9	0,0	0,0	0,4	315	3	2	0,0	97,0	2,0	0,0	1,0
505	5	3	0,0	0,0	84,3	4,6	11,0	401	4	3	1,0	5,3	93,7	0,0	0,0
512	5	3	0,0	0,1	82,9	0,1	17,0	404	4	1	98,1	1,9	0,0	0,0	0,0
								405	4	1	57,4	0,5	0,0	42,1	0,0
								407	4	1	97,6	2,4	0,0	0,0	0,0
								411	4	1	98,2	1,9	0,0	0,0	0,0
								508	5	2	0,0	99,9	0,0	0,1	0,0
								509	5	2	10,9	66,6	0,0	0,0	22,5
Error de Clasificación: 16,5%							Error de Clasificación: 20,7%								

Sin embargo, a diferencia de los resultados obtenidos mediante análisis discriminante canónico, en este caso los errores de clasificación de los modelos desarrollados empleando como variables los datos espectrales NIR fueron ligeramente superiores a los obtenidos a partir de los valores predichos de ácidos grasos, independientemente de la forma de presentación de muestra considerada.

Así, en el caso de grasa fundida, el error de clasificación fue de 1,7% al emplear como variables en el modelo los datos predichos de ácidos grasos y del 3,3% al considerar los datos espectrales NIR. De igual modo, al considerar los modelos generados en tejido adiposo, el error de clasificación obtenido al emplear como variables los valores predichos de ácidos grasos fue del 16,5%, mientras que al considerar los datos espectrales NIR fue del 20,7%.

Si se observan con más detalle la matriz de clasificación en los modelos generados a partir de los datos espectrales NIR en grasa fundida (Tabla 4.4.2), las diferencias en estos errores se deben a un mayor error parcial de clasificación de los lotes 1 y 2. Así, 2 muestras del lote 1 son clasificadas como pertenecientes al lote 2 (107 y 138) y una muestra del lote 2 (227) es clasificada como perteneciente al lote 1.

Teniendo en cuenta la semejanza entre el tipo de alimentación empleado en estos lotes y la variabilidad animal existente, estos errores de clasificación se pueden considerar asumibles por el modelo, máxima teniendo en cuenta que la probabilidad de clasificación de algunas de estas muestras (107 y 227) es similar tanto para el lote 1 como para el lote 2 (Prob. Lote 1 = 47,7% vs. Prob. Lote 2 = 52,2% para la muestra 107).

En el caso de los modelos desarrollados para tejido adiposo, la principal diferencia entre los errores de clasificación totales de los modelos generados a partir de valores de ácidos grasos y datos espectrales NIR se debió, básicamente, al mayor error parcial del lote 1 en el modelo desarrollado mediante datos espectrales *per se*. Así, como se observa en la Tabla 4.4.2, en este modelo el error parcial del lote 1 fue de un 24,4% (4,9% + 19,5%) mientras que para el modelo generado a partir de valores de ácidos grasos, el error fue de un 7,3% (4,9% + 2,4%).

Esta diferencia vino motivada por una mayor clasificación errónea de muestras del lote 1 en el lote 4, ya observada en los resultados del análisis discriminante

canónico, si bien se ha de tener en cuenta que la mayoría de estas muestras presentaron una probabilidad de pertenecer al Lote 1, y por tanto de ser correctamente clasificadas, superior al 30% (Tabla 4.4.3).

Además, tras comparar los resultados de los modelos desarrollados para tejido adiposo se puede apreciar que algunas de las muestras mal clasificadas lo son tanto por el modelo generado a partir de valores de ácidos grasos como de datos espectrales NIR (Tabla 4.4.3). Este es el caso de las muestras 201, 202, 203, 204, 312, 315 y 411, destacando que la clasificación de estas muestras mediante ambos modelos fue similar. Ello implica que, independientemente de las variables consideradas para el desarrollo de los modelos discriminantes cuadráticos, dichas muestras son clasificadas en otros lotes diferentes a los de su origen. Por ello, y teniendo en cuenta que estas muestras no fueron mal clasificadas por los modelos generados para grasa fundida, habría que sospechar en un posible error en la toma de espectros de tejido adiposo, dado que los valores de ácidos grasos provienen de predicciones realizadas a partir de espectros NIR.

Otro aspecto a resaltar de los resultados obtenidos mediante los modelos para tejido adiposo es el error parcial de clasificación existente en el lote 5, donde un 15,4% de las muestras son clasificadas erróneamente en lotes que han consumido algo de bellota durante su periodo de cebo. Este resultado supone un claro error de clasificación del modelo ya que los animales pertenecientes al lote 5 tuvieron un régimen alimenticio durante el cebo basado exclusivamente en pienso.

4.4.2.3 *Análisis discriminante lineal*

A pesar de que el test de homogeneidad de las matrices de covarianza recomendó, en todos los modelos estudiados, el desarrollo de un análisis discriminante cuadrático, se realizó un análisis discriminante lineal para poder comparar los resultados obtenidos con ambos algoritmos discriminantes y asimismo porque el análisis discriminante lineal ha sido el tradicionalmente empleado en otros trabajos de clasificación de grasas y aceites (Hervás *et al.*, 1994; De Pedro *et al.*, 1995; Indahl *et al.*, 1999; Hourant *et al.*, 2000).

En la Tabla 4.4.4 se muestran las matrices de clasificación resultantes de los modelos desarrollados mediante análisis discriminante lineal, a partir de los valores predichos de ácidos grasos y de datos espectrales NIR en grasa fundida y tejido adiposo.

Tabla 4.4.4. Resultados de modelos desarrollados mediante análisis discriminante lineal con los valores predichos de ácidos grasos y con datos espectrales NIR.

Ácidos grasos en grasa fundida						Datos NIRS en grasa fundida					
CLASIFICADO EN LOTE						CLASIFICADO EN LOTE					
1 2 3 4 5						1 2 3 4 5					
LOTE ORIGEN	1	40 97,6%	-	-	1 2,4%	-	41 100,0%	-	-	-	-
	2	-	36 100,0%	-	-	-	-	36 100,0%	-	-	
	3	-	-	20 100,0%	-	-	-	-	20 100,0%	-	
	4	1 9,1%	-	-	10 90,9%	-	-	-	-	11 100,0%	-
	5	-	-	-	-	13 100,0%	-	-	-	-	13 100,0%
Error de Clasificación: 1,7%						Error de Clasificación: 0,0%					
Ácidos grasos en tejido adiposo						Datos NIRS en tejido adiposo					
CLASIFICADO EN LOTE						CLASIFICADO EN LOTE					
1 2 3 4 5						1 2 3 4 5					
LOTE ORIGEN	1	33 80,5%	8 19,5%	-	-	-	29 70,7%	-	-	12 29,3%	-
	2	3 8,3%	22 61,1%	4 11,1%	7 19,5%	-	-	26 72,2%	9 25,0%	1 2,8%	-
	3	-	1 5,0%	19 95,0%	-	-	-	-	20 100,0%	-	-
	4	1 9,1%	-	1 9,1%	9 81,8%	-	1 9,1%	-	1 9,1%	6 81,8%	-
	5	-	-	2 15,4%	-	11 84,6%	-	-	2 15,4%	-	11 84,6%
Error de Clasificación: 22,3%						Error de Clasificación: 21,5%					

En estos modelos, la selección del número de componentes principales se realizó siguiendo el mismo criterio que para el análisis discriminante cuadrático. Así, en los modelos generados para muestras de grasa fundida, se seleccionaron 4 componentes principales al emplear como variables los valores de ácidos grasos y 7 al considerar los datos espectrales NIR. Para los modelos desarrollados a partir de muestras de tejido adiposo, el número de componentes principales seleccionadas fue de 8 al emplear los valores de ácidos grasos como variables y de 10 al considerar los datos espectrales NIR.

Comparando estos valores de componentes principales con los indicados en el análisis discriminante cuadrático puede apreciarse que, en este caso, el número de componentes principales seleccionadas fue ligeramente superior. Es decir, fue necesario emplear un mayor número de componentes principales para obtener el valor mínimo del error de clasificación mediante validación cruzada.

De igual modo que en los resultados obtenidos mediante análisis discriminante canónico y cuadrático, los modelos generados mediante análisis discriminante lineal para grasa fundida permiten diferenciar mejor los distintos lotes del colectivo de aprendizaje que aquellos desarrollados para muestras de tejido adiposo. Así, el error de clasificación, estimado mediante validación cruzada completa, fue inferior al 2% en el caso de los modelos generados en grasa fundida mientras que los modelos desarrollados en tejido adiposo, presentó un valor superior al 20% (Tabla 4.4.4).

Sin embargo, a diferencia de los resultados obtenidos mediante análisis discriminante cuadrático, los errores de clasificación de los modelos desarrollados empleando como variables los datos espectrales NIR fueron ligeramente inferiores a los obtenidos a partir de los valores predichos de ácidos grasos, independientemente de la forma de presentación de muestra considerada. Estos resultados coinciden con los obtenidos mediante análisis discriminante canónico.

Una posible causa de los diferentes resultados obtenidos en el análisis discriminante cuadrático es la descrita por Næs *et al.* (2002). Según estos autores, el análisis discriminante cuadrático genera modelos más inestables que el análisis discriminante lineal por lo que los errores pueden llegar a ser superiores.

En la Tabla 4.4.5 se indican las muestras mal clasificadas por los modelos desarrollados mediante análisis discriminante lineal junto con el valor de probabilidad (%) de pertenencia de cada muestra a cada uno de los lotes (categorías).

En las matrices de clasificación descritas en la Tabla 4.4.4, puede apreciarse que el modelo generado a partir de datos NIRS obtenidos de grasa fundida, el error de clasificación mediante validación cruzada fue del 0%, resultando todas las muestras correctamente clasificadas en sus lotes de origen. Sin embargo, el modelo generado a

partir de valores predichos de ácidos grasos para esta misma forma de presentación de muestra fue idéntico al observado para análisis discriminante cuadrático.

Tabla 4.4.5. Muestras mal clasificadas por modelos generados mediante análisis discriminante lineal a partir de valores predichos de ácidos grasos y datos espectrales NIR.

Ácidos grasos en grasa fundida							Datos NIRS en grasa fundida											
Muest.	Lote	Clasif.	PROBABILIDAD CLASIFICACIÓN LOTE (%)					Muest.	Lote	Clasif.	PROBABILIDAD CLASIFICACIÓN LOTE							
			1	2	3	4	5				1	2	3	4	5			
132	1	4	9,1	0,0	0,0	91,0	0,0											
409	4	1	97,0	0,0	1,0	2,0	0,0											
Error de Clasificación: 1,7%							Error de Clasificación: 0,0%											
Ácidos grasos en tejido adiposo							Datos NIRS en tejido adiposo											
Muest.	Lote	Clasif.	PROBABILIDAD CLASIFICACIÓN LOTE					Muest.	Lote	Clasif.	PROBABILIDAD CLASIFICACIÓN LOTE							
			1	2	3	4	5				1	2	3	4	5			
107	1	2	4,6	66,9	9,3	19,3	0,0	107	1	4	34,7	0,0	0,0	65,3	0,0			
114	1	2	5,0	40,8	16,8	37,4	0,0	125	1	4	4,2	42,4	1,0	52,5	0,0			
125	1	2	15,3	81,9	0,2	2,6	0,0	126	1	4	38,1	0,0	0,0	61,9	0,0			
131	1	2	46,6	47,8	0,0	5,5	0,0	129	1	4	14,7	0,0	0,0	85,3	0,0			
132	1	2	2,3	85,0	0,8	12,0	0,0	131	1	4	28,2	0,0	0,0	71,8	0,0			
138	1	2	42,0	50,8	2,4	4,8	0,0	132	1	4	12,2	0,0	0,0	87,8	0,0			
139	1	2	28,9	66,9	0,7	3,5	0,0	133	1	4	31,3	0,0	0,0	68,7	0,0			
150	1	2	28,2	69,4	0,6	1,9	0,0	134	1	4	34,3	0,0	0,0	65,7	0,0			
201	2	4	34,0	14,9	0,1	51,1	0,0	136	1	4	10,6	0,0	0,0	89,4	0,0			
205	2	3	5,9	23,2	69,7	1,2	0,0	138	1	4	20,0	0,0	0,0	80,0	0,0			
208	2	4	1,6	21,3	1,1	76,0	0,0	139	1	4	24,5	4,6	0,3	70,6	0,0			
210	2	4	1,3	28,1	0,2	70,4	0,0	141	1	4	45,1	0,0	0,0	54,9	0,0			
211	2	4	0,8	0,2	0,0	99,0	0,0	218	2	3	0,0	3,4	96,6	0,1	0,0			
215	2	1	84,2	13,0	0,0	2,8	0,0	223	2	3	0,0	15,7	84,3	0,0	0,0			
217	2	4	0,2	14,3	1,2	84,4	0,0	226	2	3	0,0	8,6	91,3	0,1	0,0			
218	2	3	0,0	0,1	99,9	0,0	0,0	228	2	3	0,0	24,8	75,3	0,0	0,0			
223	2	1	55,3	42,6	2,0	0,0	0,0	230	2	3	0,0	14,9	85,1	0,0	0,0			
226	2	3	3,1	31,4	64,8	0,7	0,0	232	2	4	38,5	8,0	1,0	52,6	0,0			
232	2	1	76,0	17,6	0,0	6,4	0,0	233	2	3	0,0	2,1	97,7	0,2	0,0			
235	2	4	0,6	40,4	13,4	45,7	0,0	234	2	3	0,0	43,0	57,0	0,0	0,0			
236	2	3	0,0	14,7	82,0	3,2	0,1	235	2	3	0,0	20,0	79,9	0,1	0,0			
237	2	4	0,0	33,4	31,7	34,9	0,0	237	2	3	0,0	28,7	71,4	0,0	0,0			
312	3	2	1,5	57,6	39,6	1,3	0,0	401	4	3	0,6	47,1	49,3	3,0	0,0			
401	4	3	0,4	26,5	42,1	30,8	0,3	407	4	1	81,8	0,0	0,0	18,2	0,0			
409	4	1	36,6	36,1	0,5	26,8	0,0	505	5	3	0,0	0,0	88,3	0,0	11,7			
505	5	3	0,0	2,5	72,3	3,5	21,8	508	5	3	0,0	0,2	76,7	0,0	23,1			
512	5	3	0,0	1,8	90,4	2,2	5,6											
Error de Clasificación: 22,3%							Error de Clasificación: 21,5%											

El modelo desarrollado mediante análisis discriminante lineal a partir de datos NIRS *per se* de tejido adiposo presentó dificultades para diferenciar los lotes 1 y 4 ya que 12 muestras del lote 1 (29,3%) resultaron clasificadas en el lote 4 (Tabla 4.4.5).

Sin embargo, al igual que ocurrió en el análisis discriminante cuadrático, la mitad de dichas muestras (107, 126, 131, 133, 134 y 141) presentaron una alta probabilidad (superior al 30%) de ser correctamente clasificadas.

Este mayor error parcial de clasificación del lote 1 no fue tan elevado (19,5%) en el caso del modelo generado a partir de valores predichos de ácidos grasos en tejido adiposo, no resultando las muestras clasificadas incorrectamente en el lote 4 sino en el lote 2, con un tipo de alimentación más similar al del lote 1. Sin embargo, 1 muestra del lote 4 (409) fue clasificada de nuevo incorrectamente en el lote 1. Excepto en el lote 1 donde el error parcial fue superior, en el resto de los lotes, los errores parciales de clasificación del modelo desarrollado a partir de valores de ácidos grasos en tejido adiposo fueron iguales o superiores a los obtenidos en el modelo que empleó datos espectrales NIR en la misma forma de presentación de muestra.

Tras comparar los resultados de los modelos desarrollados para tejido adiposo se puede apreciar que algunas de las muestras mal clasificadas lo son, al igual que en análisis discriminante cuadrático, tanto por el modelo generado a partir de valores de ácidos grasos como de datos espectrales NIR (Tabla 4.4.5). Este fue el caso de las muestras 107, 125, 131, 132, 138, 139, 218, 223, 226, 232, 235, 237 y 401, las cuales fueron clasificadas por ambos modelos en lotes de similares características en cuanto a alimentación.

Dado que, a excepción de la muestra 132, ninguna de ellas resultó mal clasificadas por los modelos generados para grasa fundida, habría que sospechar de nuevo en un posible error en la toma de espectros de tejido adiposo. El caso de la muestra 132 del lote 1 es diferente pues es clasificada incorrectamente por 3 modelos de los 4 generados (es clasificada por 2 modelos en el lote 4 y por 1 en el lote 2). Ello significaría que dicha muestra presenta alguna particularidad como para ser clasificada mediante análisis discriminante lineal incorrectamente. Además, teniendo en cuenta que la misma muestra resultó clasificada incorrectamente mediante 2 modelos de análisis discriminante cuadrático, cabría pensar en un posible error en cuanto a la información de campo suministrada del animal del cual se tomó dicha muestra.

Un aspecto a resaltar de los resultados obtenidos mediante análisis discriminante lineal en muestras de tejido adiposo es el error parcial de clasificación

existente en el lote 5, resultado similar al ya comentado al realizar un análisis discriminante cuadrático.

Teniendo en cuenta los resultados obtenidos en los diferentes tipos de análisis discriminante empleados, es posible concluir sobre la mayor capacidad de clasificación de aquellos modelos quimiométricos multivariantes que usan como variable la información espectral NIR de grasa fundida frente a los que emplean el espectro de tejido adiposo.

Resulta difícil evaluar esta inferior capacidad discriminatoria de los modelos desarrollados en tejido adiposo si bien se pueden emitir algunas hipótesis que ayuden a continuar los trabajos de desarrollo de modelos cualitativos (también útiles para modelos cuantitativos) en esta forma de análisis. Entre estas hipótesis estaría el mayor error de predicción de ácidos grasos en tejido adiposo que en grasa fundida, para los modelos generados a partir de la composición en ácidos grasos predichos por NIRS y el mayor error en la obtención de la medida espectroscópica NIR en tejido adiposo que en grasa fundida, para los modelos desarrollados a partir de datos espectrales *per se*. En definitiva, varias dificultades en el establecimiento de modelos NIRS en tejido adiposo, independientemente de las variables empleadas, lo que obliga a profundizar en el estudio de los modelos cualitativos (de igual modo para los cuantitativos) a partir de este tipo de información espectral, incrementando el número de muestras, disminuyendo los errores de medida espectral, profundizando en el rango de longitudes de onda a considerar, etc.

A pesar de ello, aquellos modelos discriminantes obtenidos a partir de datos espectrales NIR diferencian de forma más clara y, en general, como menor error los lotes con diferentes regímenes alimenticios de aquellos modelos en los que se emplea los valores predichos de ácidos grasos como variables.

Esto puede ser debido a que espectro NIR representa la “huella digital” de cada individuo, es decir, incorpora información de todos los analitos presentes en una muestra de grasa y no únicamente, de su composición en ácidos grasos mayoritarios y minoritarios.

4.4.3 DESARROLLO DE MODELOS CUALITATIVOS MEDIANTE ANÁLISIS SIMCA

Como ya se indicó en el apartado 3.4.2.2, el análisis SIMCA difiere del análisis discriminante expuesto en el apartado anterior en el hecho de que cada clase es modelada por separado, clasificándose las muestras en función de sus distancias a cada modelo.

Así, en el análisis SIMCA se modela independiente cada una de las clases mediante un análisis de componentes principales, determinando con ello el centroide y la dispersión de los valores para cada clase, quedando definida el modelo independiente de cada clase por el número de componentes principales que se hayan considerado.

El modelo independiente obtenido para cada clase es definido en base a dos distancias, las cuales se denominan distancia de las muestras al centro del modelo y distancia de la muestra al modelo. En el caso del programa empleado (Unscrambler), estas distancias se denominan *Leverage* y *S* respectivamente (Martens y Næs, 1989; Esbensen *et al.*, 1994). Para que una muestra sea clasificada como perteneciente a una determinada clase ambos estadísticos (*leverage* y *S*) han de ser inferiores a los valores máximos definidos para dichos estadísticos en cada clase.

Dado que en el análisis SIMCA se realiza un análisis de componentes principales individual en cada una de las clases, se han empleado directamente como variables los valores predichos de ácidos grasos y los datos espectrales NIR (con un tratamiento de derivada del tipo (2,5,5,1)) de cada uno de los lotes, y no se ha realizado una síntesis previa y global de la información como en los diferentes tipos de análisis discriminantes.

Para la evaluación de los modelos cualitativos generados se procedió a la clasificación en grupos de las muestras del colectivo de aprendizaje, empleando para ello los diferentes modelos desarrollados de forma independiente para cada clase. En realidad, esta etapa no consiste en una reclasificación ya que en la generación de cada uno de los modelos únicamente han intervenido las muestras de su clase.

El análisis SIMCA es un análisis de modelado mediante el cual una misma muestra puede ser asignada a una única clase, a varias o a ninguna de las clases

(apartado 3.4.2.2). Por ello, los estadísticos empleados han de ser diferentes al considerado en un análisis de clasificación propiamente dicho como es un análisis discriminante (error de clasificación). Así, los estadísticos empleados en la interpretación de resultados de un análisis SIMCA fueron el número de falsos negativos y falsos positivos. En un modelo ideal, el valor de muestras detectadas como falsos negativos o falsos positivos ha de ser cero.

En la Tabla 4.4.6 se muestran los resultados de los modelos generados mediante análisis SIMCA a partir de valores predichos de ácidos grasos y de datos espectrales NIR en grasa fundida y tejido adiposo.

Tabla 4.4.6. Resultados de modelos desarrollados mediante análisis SIMCA a partir de valores predichos de ácidos grasos y datos espectrales NIR*.

Ácidos grasos en grasa fundida				Datos NIRS en grasa fundida					
	FN	FP	Error Lote		FN	FP	Error Lote		
LOTE	1	4	1	0,0	LOTE	1	7	0	0,0
	2	3	0	0,0		2	1	0	0,0
	3	2	0	0,0		3	1	0	0,0
	4	0	2	9,1		4	0	0	0,0
	5	2	0	0,0		5	1	0	0,0
Error de Clasificación: 0,8%				Error de Clasificación: 0,0%					
Ácidos grasos en tejido adiposo				Datos NIRS en tejido adiposo					
	FN	FP	Error Lote		FN	FP	Error Lote		
LOTE	1	4	15	7,3	LOTE	1	3	15	7,3
	2	8	34	5,6		2	0	28	5,6
	3	0	2	45,0		3	2	2	35,0
	4	0	5	72,7		4	1	25	54,5
	5	0	3	7,7		5	1	3	0,0
Error de Clasificación: 19,0%				Error de Clasificación: 14,9%					

*FP: Falsos positivos; FN: Falsos negativos

Como se aprecia en dicha tabla, el número de falsos negativos obtenidos mediante los diferentes modelos generados para grasa fundida y tejido adiposo fue similar, oscilando entre 7 y 12. Ello implica que un determinado número de muestras (algo menos del 10% del total del colectivo de aprendizaje) no es asignado correctamente en su propia clase. Estas muestras detectadas como falsos negativos se observan en todos los lotes, independientemente del modelo considerado, a excepción

del lote 4 en el cual únicamente presenta una muestra como falso negativo en el modelo generado a partir de datos espectrales NIR de tejido adiposo subcutáneo. En general, el lote que presenta un mayor número de muestras como falsos negativos es el lote 1, el cual incluso detecta 7 muestras como falsos negativos en el modelo generado a partir de datos NIRS en grasa fundida.

Sin embargo, el número de muestras detectadas como falsos positivos es claramente diferente dependiendo del modelo considerado, siendo mucho más elevado en los modelos generados para tejido adiposo (Tabla 4.4.6). Ello indica que en estos modelos, un elevado número de muestras son asignadas de forma incorrecta a más de un lote, presentando por ello una baja especificidad. En los modelos generados para grasa fundida, el número de muestras detectadas como falsos positivos es muy pequeño, llegando a ser nulo para el modelo generado a partir de datos espectrales NIR. Por ello, este modelo presenta una máxima especificidad, de forma que todas las muestras que no pertenecen a una determinada clase son reconocidas como tales.

Como ya se indicó en el apartado 3.4.2.2, todo análisis de modelado puede ser empleado como análisis de clasificación. En el caso del análisis SIMCA, el procedimiento consistiría en clasificar las muestras en aquella clase en la cual presente menores valores de distancias *leverage* y *S*. A partir de estos resultados, es posible definir, al igual que en el análisis discriminante, un error de clasificación parcial (para cada lote) y global (para todo el modelo). En la Tabla 4.4.6 se han incluido tanto los valores de los errores de clasificación parciales de cada lote y el error global para los diferentes modelos generados.

Además en la Tabla 4.4.7 se indican las muestras mal clasificadas por los modelos desarrollados mediante análisis SIMCA tras asignar las muestras a aquel lote en el cual presenten menores valores de *leverage* y *S*. Así, para cada muestra, se indican los lotes a los cuales pertenecerían según sus estadísticos *leverage* y *S* y aquel lote en el cual estaría finalmente clasificado al presentar los menores valores.

Los resultados incluidos en la Tabla 4.4.6 confirman aquellos obtenidos mediante análisis discriminante de forma que los modelos desarrollados en muestra de grasa fundida presentan un menor valor en el error de clasificación global, permitiendo diferenciar mejor los distintos lotes del colectivo de calibración que aquellos

desarrollados con muestras de tejido adiposo. Así, el error de clasificación fue inferior al 1% en el caso de los modelos generados en grasa fundida mientras que los modelos desarrollados en tejido adiposo, presentó un valor superior al 15%.

Tabla 4.4.7. Muestras mal clasificadas por modelos generados mediante análisis SIMCA a partir de valores predichos de ácidos grasos y datos espectrales NIR*.

Ácidos grasos en grasa fundida						Datos NIRs en grasa fundida														
Muest.	Lote	Clasif.	LOTE ASIGNADO SEGÚN <i>LEVERAGE Y S</i>					Muest.	Lote	Clasif.	LOTE ASIGNADO SEGÚN <i>LEVERAGE Y S</i>									
			1	2	3	4	5				1	2	3	4	5					
409	4	1	⊗																	
Error de Clasificación: 0,8%						Error de Clasificación: 0,0%														
Ácidos grasos en tejido adiposo						Datos NIRs en tejido adiposo														
Muest.	Lote	Clasif.	LOTE ASIGNADO SEGÚN <i>LEVERAGE Y S</i>					Muest.	Lote	Clasif.	LOTE ASIGNADO SEGÚN <i>LEVERAGE Y S</i>									
			1	2	3	4	5				1	2	3	4	5					
125	1	2	×	⊗							×	⊗								×
138	1	2	×	⊗																×
150	1	2	×	⊗																
201	2	1	⊗								⊗	×								×
215	2	1	⊗								⊗	×								×
304	3	2		⊗		×						⊗		×						
306	3	2		⊗		×						⊗		×						
307	3	2		⊗		×						⊗		×						
308	3	2		⊗		×						⊗		×						
309	3	2		⊗		×						⊗		×						
312	3	2		⊗		×						⊗		×						
315	3	5				×														⊗
318	3	2		⊗		×						⊗		×						
351	3	2		⊗		×						⊗		×						
401	4	2		⊗								⊗			×					
402	4	2		⊗								⊗			×					
403	4	2	×	⊗								⊗			×					
404	4	2	×	⊗								⊗			×					
405	4	2	×	⊗								⊗			×					
407	4	2		⊗								⊗			×					
409	4	2	×	⊗								⊗			×					
412	4	2		⊗								⊗			×					
512	5	3				⊗														×
Error de Clasificación: 19,0%						Error de Clasificación: 14,9%														

*Clasificación: ⊗; Pertenencia a una clase: ×

Los valores más altos del error de clasificación para los modelos generados en muestras de tejido adiposo, a partir de valores predichos de ácidos grasos o datos espectrales NIR, se obtuvieron en los lotes 3 y 4, siendo dichos errores superiores al

35% (Tabla 4.4.6). Observando con más detalle las muestras mal clasificadas de dichos lotes (Tabla 4.4.7), se aprecia como estas son clasificadas en el lote 2 e incluso, en algunas ocasiones, en el lote 1, al presentar el menor valor de los estadísticos *leverage* y *S*. Sin embargo, la mayor parte de estas muestras mal clasificadas también fueron asignadas como pertenecientes a su lote de origen, si bien los valores de los estadísticos fueron superiores.

Es de destacar que, al igual que en los modelos desarrollados mediante análisis discriminante, un cierto número de muestras resultaron mal clasificadas por los modelos generados en tejido adiposo, independientemente de la variable considerada para su desarrollo (valores de ácidos grasos o datos NIRS). En este caso, fueron las muestras 125, 201, 309, 312, 318, 401, 403, 407 y 412. El hecho de no resultar mal clasificadas dichas muestras mediante los modelos generados en grasa fundida, implicaría que el origen de dicha clasificación errónea fuera debido al propio espectro NIR obtenido para esta forma de presentación de muestra.

Al igual que los resultados del análisis discriminante lineal, los errores de clasificación de los modelos desarrollados empleando como variables los datos espectrales NIR fueron ligeramente inferiores a los obtenidos a partir de los valores predichos de ácidos grasos, independientemente de la forma de presentación de muestra considerada. Este error de clasificación fue incluso del 0% (todas las muestras correctamente clasificadas en su lote) en el modelo desarrollado a partir de datos espectrales NIR en grasa fundida.

4.4.4 DESARROLLO DE MODELOS CUALITATIVOS MEDIANTE ANÁLISIS DISCRIMINANTE BASADO EN REGRESIÓN RMCP

Como se ha indicado en el apartado 3.4.2.3, el análisis discriminante basado en una regresión RMCP, a diferencia de los diferentes tipos de análisis discriminantes descrito en el apartado 3.4.2.1, es un análisis de modelado mediante el cual se generan ecuaciones discriminantes utilizando un algoritmo de regresión por mínimos cuadrados parciales ó RMCP. Se generan tantas variables discriminatorias como clases (lotes) existan en el grupo de aprendizaje de forma que cada una de ellas sintetiza información

tanto de las variables consideradas (ácidos grasos o datos NIRS) como la relativa a la variable cualitativas de referencia (grupo de clasificación, en este caso).

Mediante el paquete quimiométrico WinISI, se ha utilizado el algoritmo denominado RMCP2 utilizado para modelos de clasificación en los que se dispone de más de 2 clases a discriminar (Shenk y Westerhaus, 1995b; Næs *et al.*, 2002). A las variables discriminatorias se les asignan valores entre 1 ó 2 según pertenezca una muestra a la clase o no, respectivamente.

Una de las dificultades de este tipo de análisis estriba en establecer un valor umbral a partir del cual se considere la pertenencia de una muestra a una determinada clase. Diferentes autores (Downey, 1996 y 2000; Næs *et al.*, 2002) utilizan un valor de 1,5 como límite para establecer la pertenencia o no a un determinado grupo, de forma que si una muestra presenta un valor inferior a 1,5 para una variable discriminatoria, no pertenecería a dicha clase. Si el valor de la variable discriminatoria es superior o igual a 1,5, la muestra sería clasificada dentro de la clase correspondiente. Este fue el valor umbral utilizado en el presente estudio.

Este algoritmo de modelado realiza una síntesis de la información redundante de las variables iniciales si bien, en este caso, se emplea una regresión RMCP y no un análisis de componentes principales como en los algoritmos empleados en los apartados previos. Por ello, al igual que en el análisis SIMCA, se han empleado directamente como variables los valores predichos de ácidos grasos y los datos espectrales NIR (con un tratamiento de derivada del tipo (2,5,5,1)) de cada uno de los lotes. Para calcular el número óptimo de términos de la regresión, se utilizó una validación cruzada, seleccionando aquel número que minimizara el error típico de validación cruzada (*ETVC*).

Al igual que en el análisis SIMCA, en este tipo de análisis discriminante una misma muestra podría estar clasificada en una, en varias o en ninguna de las clases. Por ello, los estadísticos empleados van a ser el número de falsos negativos y falsos positivos.

En la Tabla 4.4.8 se muestran los resultados de los modelos generados mediante análisis discriminante basado en regresión RMCP a partir de valores predichos de ácidos grasos y de datos espectrales NIR en grasa fundida y tejido adiposo.

Tabla 4.4.8. Resultados de modelos desarrollados mediante análisis discriminante basado en regresión RMCP a partir de valores predichos de ácidos grasos y datos espectrales NIR*.

Ácidos grasos en grasa fundida				Datos NIRS en grasa fundida				
	FN	FP	Error Lote		FN	FP	Error Lote	
LOTE	1	0	7	0,0	1	0	0	0,0
	2	0	0	0,0	2	0	2	0,0
	3	3	3	5,0	3	0	0	0,0
	4	7	1	45,5	4	3	1	0,0
	5	0	0	7,7	5	0	0	0,0
Error de Clasificación: 5,8%				Error de Clasificación: 0,0%				
Ácidos grasos en tejido adiposo				Datos NIRS en tejido adiposo				
	FN	FP	Error Lote		FN	FP	Error Lote	
LOTE	1	8	6	17,1	1	2	1	2,4
	2	24	8	41,7	2	4	2	2,8
	3	11	3	10,0	3	0	0	0,0
	4	10	1	90,9	4	5	1	36,4
	5	2	0	0,0	5	0	0	0,0
Error de Clasificación: 28,9%				Error de Clasificación: 5,0%				

*FP: Falsos positivos; FN: Falsos negativos

Como se observa en dicha tabla, el número de falsos negativos obtenidos fue muy diferente dependiendo del modelo generado. Así, en el caso del modelo desarrollado para grasa fundida a partir de datos espectrales NIR, el número de falsos positivos fue pequeño (únicamente 3 en el lote 4). En los modelos generados en grasa fundida a partir de valores de ácidos grasos y en tejido adiposo a partir de datos espectrales NIR, el número de falsos positivos fue similar (aproximadamente 10) distribuidos entre los lotes 1 a 4. Estos resultados fueron similares a los observados en los modelos desarrollados mediante análisis SIMCA.

Sin embargo, en el modelo generado en tejido adiposo a partir de datos de composición en ácidos grasos, el número total de falsos negativos fue muy elevado (superior a 50), detectándose en todos los lotes (fundamentalmente en el lote 2). Esto supone que casi el 50% de las muestras del colectivo de aprendizaje no son asignadas correctamente a su propio lote. Este resultado es anormal pues no hay ningún motivo a priori que pueda provocar este valor tan elevado de falsos negativos, teniendo en cuenta que el resultado obtenido para este mismo modelo mediante análisis SIMCA fue muy bajo y del mismo orden al del resto de los modelos.

En cuanto al número observado de falsos positivos o muestras asignadas incorrectamente como pertenecientes a un determinado lote, indicar que no fue muy elevado en ninguno de los modelos si bien los valores fueron diferentes dependiendo de las variables consideradas (Tabla 4.4.8).

Así, en los modelos generados a partir de valores predichos de ácidos grasos, el número de falsos positivos osciló entre 11 y 18 dependiendo de la forma de presentación de muestra empleada. En los modelos desarrollados a partir de valores espectrales NIR, este número de falsos positivos fue menor, siendo tanto en el caso de muestra de grasa fundida como de tejido adiposo, inferior a 4. Por ello, al emplear este análisis discriminante, los modelos generados a partir de valores espectrales NIR presentan una mayor especificidad que los modelos desarrollados a partir de valores de ácidos grasos, de forma que todas las muestras que no pertenecen a una determinada clase son rechazadas por su modelo.

De igual modo que el análisis SIMCA, el análisis discriminante basado en regresión RMCP puede ser empleado como método de clasificación. Para ello, cada muestra ha de ser clasificada en aquel lote que presente un mayor valor de variable discriminatoria, pudiendo calcular así un error de clasificación, tanto para cada una de los lotes como el global del modelo. En la Tabla 4.4.8 se incluyen dichos errores de clasificación parciales y globales para los diferentes modelos generados.

Además, en la Tabla 4.4.9 se indican las muestras mal clasificadas por los modelos desarrollados mediante este análisis discriminante, tras clasificar las muestras en aquel lote donde presente un mayor valor de la variable discriminatoria. Así, para cada muestra, se indican los lotes a los cuales pertenecerían al presentar un valor de variable discriminatoria superior a 1,5 y a aquel lote en el cual estaría finalmente clasificada, al presentar el mayor valor de las diferentes variables discriminatorias.

De nuevo, los resultados indicados en la Tabla 4.4.9 confirman los obtenidos mediante análisis discriminante y análisis SIMCA, al presentar los modelos desarrollados en grasa fundida un menor valor en el error de clasificación global que aquellos generados en muestras de tejido adiposo, considerando en dicha comparación un mismo tipo de variables.

Tabla 4.4.9. Muestras mal clasificadas por modelos generados por análisis discriminante basado en regresión RMCP a partir de valores predichos de ácidos grasos y datos espectrales NIR.

Ácidos grasos en grasa fundida							Datos NIRS en grasa fundida								
Muest.	Lote	Clasif.	VALOR VARIABLE DISCRIMINATORIA					Muest.	Lote	Clasif.	VALOR VARIABLE DISCRIMINATORIA				
			V 1	V 2	V 3	V 4	V 5				V 1	V 2	V 3	V 4	V 5
316	3	1	1,50	1,14	1,47	0,73	1,16								
401	4	1	1,55	0,93	0,98	1,31	1,23								
403	4	1	1,55	1,07	0,87	1,23	1,27								
404	4	1	1,83	0,89	0,51	1,48	1,29								
405	4	1	1,64	0,94	0,60	1,56	1,27								
409	4	1	1,41	1,05	1,21	1,27	1,07								
509	5	3	1,09	0,54	1,68	1,02	1,67								
Error de Clasificación: 5,8%							Error de Clasificación: 0,0%								
Ácidos grasos en tejido adiposo							Datos NIRS en tejido adiposo								
Muest.	Lote	Clasif.	VALOR VARIABLE DISCRIMINATORIA					Muest.	Lote	Clasif.	VALOR VARIABLE DISCRIMINATORIA				
			V 1	V 2	V 3	V 4	V 5				V 1	V 2	V 3	V 4	V 5
114	1	3	1,27	1,14	1,60	1,23	0,76	124	1	4	1,49	1,29	0,65	1,52	1,05
125	1	2	1,44	1,49	1,23	0,97	0,87	218	2	3	1,25	1,16	1,39	0,98	1,22
127	1	2	1,49	1,60	1,20	1,02	0,70	401	4	1	1,30	1,18	1,26	1,13	1,14
128	1	2	1,39	1,68	0,93	1,26	0,73	402	4	1	1,39	1,10	0,81	1,35	1,36
131	1	2	1,40	1,59	0,98	1,13	0,90	405	4	2	1,31	1,57	0,86	1,44	0,83
132	1	2	1,13	1,66	1,12	1,16	0,93	412	4	1	1,59	0,78	1,13	1,15	1,36
137	1	2	1,48	1,61	0,91	1,12	0,89								
201	2	1	1,50	1,19	1,11	1,29	0,92								
202	2	3	1,32	1,14	1,33	1,17	1,04								
203	2	1	1,33	1,23	1,33	1,01	1,10								
205	2	1	1,42	1,13	1,36	0,92	1,17								
206	2	1	1,38	1,24	1,15	1,09	1,14								
207	2	3	1,31	1,32	1,36	1,12	0,89								
208	2	1	1,35	1,22	1,13	1,24	1,06								
209	2	3	1,28	1,30	1,35	0,95	1,13								
210	2	1	1,33	1,33	1,07	1,22	1,05								
211	2	1	1,75	1,10	0,32	1,51	1,32								
215	2	1	1,65	1,13	1,25	1,00	0,97								
216	2	1	1,42	1,32	0,96	1,15	1,16								
217	2	1	1,29	1,13	1,26	1,29	1,03								
218	2	3	0,76	1,23	1,67	1,18	1,16								
223	2	1	1,63	1,53	0,78	0,77	1,30								
232	2	1	1,63	1,51	0,70	1,26	0,91								
312	3	1	1,43	1,28	1,38	0,90	1,01								
315	3	5	0,93	1,36	1,30	1,01	1,40								
401	4	1	1,40	1,14	1,10	1,05	1,31								
402	4	2	1,12	1,47	0,88	1,24	1,30								
403	4	2	1,18	1,52	0,98	1,24	1,08								
404	4	2	1,18	1,61	0,95	1,36	0,90								
405	4	2	1,31	1,41	0,92	1,33	1,04								
406	4	2	1,01	1,49	1,12	1,34	1,04								
407	4	1	1,56	1,22	0,92	1,24	1,08								
409	4	1	1,50	1,23	1,02	1,18	1,08								
411	4	1	1,38	1,34	0,92	1,29	1,07								
412	4	2	1,13	1,53	1,02	1,20	1,13								
Error de Clasificación: 28,9%							Error de Clasificación: 5,0%								

Sin embargo, en estos resultados se aprecia que el error de clasificación obtenido en tejido adiposo a partir de datos espectrales NIR es anormalmente bajo (error = 5%) comparado con los resultados obtenidos mediante otros algoritmos, llegando a ser incluso del mismo orden que el error del modelo generado en grasa fundida a partir de valores de ácidos grasos (error = 5,8%). Este resultado anormalmente bajo podría deberse a una sobre estimación de dicho modelo generado a partir de datos NIRS en tejido adiposo.

Los valores más altos del error de clasificación parcial para todos los modelos generados se obtuvo en el lote 4, siendo dichos errores superiores al 35% en todos los modelos a excepción de aquel desarrollado a partir de los datos NIRS en grasa fundida donde su valor fue 0% (Tabla 4.4.8). En concreto, las muestras mal clasificadas de dicho lote fueron asignadas, de forma general, en los lotes 1 y 2, hecho ya observado en los otros algoritmos empleados.

De nuevo, un cierto número de muestras resultaron mal clasificadas por los modelos generados en tejido adiposo, independientemente de la variable considerada para su desarrollo (valores de ácidos grasos o datos NIRS). En este caso, fueron las muestras 218, 401, 402, 405 y 412, ya destacadas en los resultados del análisis SIMCA (Tabla 4.4.9). Sin embargo, a diferencia de los otros algoritmos, algunas de estas muestras (401 y 405) también fueron mal clasificadas por el modelo generado a partir de ácidos grasos en grasa fundida, por lo que habría que sospechar de la presencia de alguna particularidad en ellas que les provoca su incorrecta clasificación por 3 de los 4 modelos generados.

Una muestra que merece una mención especial es la 409, perteneciente al lote 4 pero que resultó clasificada en el lote 1 por todos los modelos desarrollados con ácidos grasos en grasa fundida mediante los diferentes algoritmos matemáticos considerados. Si se tiene en cuenta la composición de los ácidos grasos mayoritarios de esta muestra, predicha en grasa fundida (C16:0 = 19,6%; C18:0 = 8,6%; C18:1 = 55,7%; C18:2 = 9,8%), se puede apreciar que dichos porcentajes son intermedios entre los valores de composición del lote 1 y del lote 4 (Tabla 4.4.1). Además, si se compara su composición con la del resto de muestras de su lote se observa que la muestra 409 presenta los valores más altos de C18:1 y los más bajos de ácidos grasos saturados

C16:0 y C18:0. Es decir, sería la muestra de este lote que más se asemejaría a los valores medios de composición de los lotes 1 y 2.

De forma general, esta muestra también fue clasificada incorrectamente en los lotes 1 ó 2 en los modelos generados a partir de ácidos grasos en tejido adiposo, presentando un comportamiento similar al anteriormente descrito en cuanto a su composición en ácidos grasos predichos en tejido adiposo. Por todo ello, cabría pensar que la incorrecta clasificación de dicha muestra por los modelos generados a partir de valores de ácidos grasos en realidad responde a su composición en ácidos grasos mayoritarios más parecida a los valores medios de los lotes 1 y 2.

En cambio, dicha muestra resultó correctamente clasificada en su lote por todos los modelos generados a partir de datos espectrales NIR, independientemente de la forma de presentación de muestra o del algoritmo matemático empleado en la generación de los modelos.

Por todo ello, se podría concluir que, de nuevo, los modelos discriminantes generados mediante regresión RMCP a partir de valores de ácidos grasos o datos espectrales NIR de muestras de grasa fundida presentaron menor error que aquellos modelos desarrollados a partir de información de muestras de tejido adiposo. Además, para una misma forma de presentación de muestra, los modelos discriminantes obtenidos a partir de datos espectrales NIR presentaron un error ligeramente inferior al obtenido en los modelos generados con valores predichos de ácidos grasos.

4.4.5 SELECCIÓN Y VALIDACIÓN DE MODELOS

Tras los resultados obtenidos mediante los diferentes algoritmos de clasificación empleados anteriormente, se podría concluir que todos los algoritmos de clasificación parecen presentar un error de clasificación similar, para un mismo tipo de variables y forma de presentación de muestra.

En la mayor parte de estos resultados, se obtuvieron menores errores de clasificación en aquellos modelos multivariantes obtenidos a partir de datos espectrales NIR *per se*, ya sea en grasa fundida o en tejido adiposo. Estos resultados confirman los obtenidos por Hervás *et al.* (1994) y De Pedro *et al.* (1995) en 3 lotes de animales bajo

condiciones experimentales controladas, si bien dichos autores únicamente consideraron es su trabajo la forma de presentación de grasa fundida. Al mismo tiempo, los modelos generados en esta Tesis a partir de datos NIR obtenidos en grasa fundida presentaron un menor error de clasificación que los descritos por González (1997).

No se han recogido en la bibliografía consultada modelos de clasificación para el régimen alimenticio de cerdos Ibéricos a partir de datos espectrales NIR en tejido adiposo. La única aproximación fue la realizada por González (1997) a partir de espectros NIR en grasa subcutánea y perirrenal picada. Sin embargo, los resultados obtenidos en dicho trabajo presentaron unos elevados errores de clasificación en comparación con los obtenidos en esta Tesis Doctoral que incluso motivaron el no recomendar esta forma de presentación de muestra.

Los resultados obtenidos en este trabajo mediante los diferentes métodos de clasificación considerado y, en particular, los obtenidos mediante análisis discriminantes mediante regresión RMCP, confirmarían la viabilidad del análisis en esta forma de presentación de muestra para evaluar el régimen alimenticio de las canales de cerdos Ibéricos. Sin embargo, se ha de seguir profundizando en esta línea de trabajo, con la finalidad de minimizar o eliminar aquellos factores (p.ej.: baja repetibilidad espectral NIR) que afectan a los elevados errores de los modelos al compararlos con el obtenidos en grasa fundida.

Como se ha observado en los anteriores apartados, los modelos generados a partir de la composición en los 11 ácidos grasos de la grasa fundida presentaron errores de clasificación ligeramente superiores a los descritos a partir de datos espectrales NIR. Sin embargo, estos errores de clasificación fueron mucho menores que los obtenidos al clasificar los lotes en función de los criterios analíticos incluidos en el Contrato Tipo Homologado (únicamente 4 ácidos grasos).

Por todo ello, se confirma lo expuesto por De Pedro (1995 y 2000) acerca de la recomendación de emplear modelos desarrollados mediante métodos de análisis multivariante, como los expuestos en esta Tesis, para la clasificación de cerdos Ibéricos en función de su alimentación. El uso de forma individual e independiente de los valores de composición de los ácidos grasos mayoritarios únicamente conduciría a errores de clasificación mucho más elevados.

Dado que ningún algoritmo de clasificación presentó resultados claramente diferentes al resto, se consideró únicamente los modelos generados mediante análisis discriminante basado en regresión RMCP en grasa fundida para realizar la etapa de evaluación (validación). La selección de dichos modelos se realizó dada la ausencia de etapas previas de importación de datos en el programa estadístico, su facilidad de manejo y la claridad en la exposición de los resultados obtenidos con los modelos.

En la Tabla 4.4.10 se incluyen los resultados de validación de los colectivos descritos en el apartado 3.4.1.1.

Tabla 4.4.10. Matriz de clasificación del colectivo de evaluación mediante modelos discriminantes basados en regresión RMCP.

		CLASIFICADO EN LOTE				
		1	2	3	4	5
LOTE ORIGEN	A	27 100%				
	B		26 57,8%	19 43,2%		
	C			20 33,3%	39 65,0%	1 1,7%

Teniendo en cuenta la información de campo, el lote A se asemejaría más al lote 1, debido a que, en ambos casos, los animales estuvieron alimentados en montanera sin ningún aporte de pienso durante un varios meses. El modelo clasificaría todos los animales del lote A como pertenecientes a la categoría definida por el lote 1.

En el caso del lote B, 26 animales (57,8%) fueron clasificados dentro de la categoría definida por el lote 2 y 19 (43,2%) en la del lote 3. En este caso, la clasificación es de nuevo correcta ya que las categorías definidas por los lotes 2 y 3 se corresponderían con regímenes alimenticios en montanera durante un máximo de 3 meses y con un cierto aporte de pienso en la dieta.

El lote C resultó clasificado de forma mayoritaria en la clase definida por el lote 4. Sin embargo, un número importante de animales (20) fueron asignados a la clase definida por el lote 3 e incluso uno (1,7%) en la del lote 5. Esta clasificación sería correcta para el lote C ya que, a pesar de su largo periodo en montanera, presentó una lenta reposición que obligó a la incorporación en la dieta de un complemento de pienso

durante el último mes. Sin embargo, el aporte de pienso no fue tan elevado como para que algún animal fuera clasificado dentro del lote 5. Por ello, dicha muestra ha de considerarse como erróneamente clasificada.

4.5 CLONACIÓN / ESTANDARIZACIÓN DE INSTRUMENTOS Y TRANSFERENCIA DE CALIBRACIONES

Para mostrar definitivamente el potencial real de implantación de la tecnología NIRS en laboratorios e industrias del sector del cerdo Ibérico es imprescindible el mostrar la existencia de procedimientos de estandarización y/o clonación de instrumentos NIRS y el poder transferir las ecuaciones de calibración NIRS para grasa sin pérdida significativa de capacidad predictiva de dichas ecuaciones.

En este apartado se abordará el estudio de los procesos de estandarización entre equipos y posterior transferencia de ecuaciones entre instrumentos NIRS ubicados en el mismo laboratorios o en laboratorios diferentes.

4.5.1 INSTRUMENTOS NIRS UBICADOS EN UN MISMO LABORATORIO

4.5.1.1 Ajuste espectral

Antes de obtener en cada uno de los instrumentos (Master y Satélite) los espectros de las muestras empleadas en la estandarización, se realizó una evaluación previa de cada equipo mediante la medida de la repetibilidad fotométrica o ruido, precisión de la longitud de onda y respuesta del instrumento.

Como se muestra en la Tabla 4.5.1, los resultados obtenidos tras esta evaluación indican que dichos parámetros se encontraban dentro de los valores

recomendados por el fabricante para considerar que los instrumentos estaban en un estado adecuado para la obtención de espectros.

Tabla 4.5.1. Diagnósticos de los instrumentos Master y Satélite.

	Master (módulo giro)	Satélite (módulo transporte; auto ganancia)
Repetibilidad Fotométrica	14	15
Precisión Longitud Onda	0,21	0,14
Respuesta Instrumento	55.108	40.720

Así, la repetibilidad fotométrica ha de ser inferior a 20, la precisión de la longitud de onda menor que 0,30 y el valor máximo de la ganancia de los detectores debe oscilar entre 55.000 y 58.000 para instrumentos sin sistema de auto ganancia y entre 32.000 y 65.500 para instrumentos con sistema de auto ganancia, valores que se cumplían en ambos instrumentos.

Uno de los criterios para evaluar una estandarización entre instrumentos NIRS es que el valor medio estimado de $RMS(C)$ entre los espectros de 2 submuestras de varias muestras, analizada una de ellas en el instrumento Satélite y otra en el Master, ha de ser igual o incluso inferior al valor de $RMS(C)$ entre los espectros de 2 submuestras de varias muestras analizadas en el Master (Dardenne *et al.*, 1992).

La Tabla 4.5.2 muestra los valores medios de $RMS(C)$ entre los espectros de 2 submuestras de 10 muestras obtenidas en el instrumento Master y entre instrumentos antes y después de estandarizar con 4 ficheros de estandarización diferentes ($STD_{Set A}$, $STD_{Set B}$, $STD_{Set C}$ y $STD_{Set D}$).

Tabla 4.5.2. $RMS(C)$ entre espectros de 2 submuestras de 10 muestras antes y después de estandarizar con 4 ficheros de estandarización diferentes.

Espectros comparados	$RMS(C)$
Master vs. Master	3.506
Master vs. Satélite previa STD	11.338
Master vs. Satélite tras $STD_{Set A}$	952
Master vs. Satélite tras $STD_{Set B}$	1.057
Master vs. Satélite tras $STD_{Set C}$	4.042
Master vs. Satélite tras $STD_{Set D}$	3.634

Como se observa en dicha tabla, los valores de $RMS(C)$ entre espectros obtenidos en el Master y en el Satélite tras la estandarización con $STD_{Set A}$ y $STD_{Set B}$ (952 y 1.057 respectivamente) fueron inferiores al valor estimado para submuestras analizadas ambas en el Master (3.506). Sin embargo, los valores de $RMS(C)$ entre espectros de submuestras analizadas en el Master y Satélite tras su estandarización con los ficheros $STD_{Set C}$ y $STD_{Set D}$ (4.042 y 3.634 respectivamente) fueron ligeramente superiores al obtenido para submuestras en el Master (3.506).

Teniendo en cuenta lo descrito por Shenk (1992), el rango de absorción espectral de los productos a analizar mediante un proceso de estandarización ha de ser similar al rango de absorción espectral de las muestras del colectivo de calibración. Por tanto, mediante el desarrollo de un proceso de estandarización a partir del mismo tipo de muestra que las incluidas en el colectivo de calibración, se obtendrán mejores resultados que al estandarizar con muestras diferentes.

Los resultados de $RMS(C)$ indicados en la Tabla 4.5.2 confirman la posibilidad de obtener un ajuste de los valores de $\text{Log}(1/R)$ obtenidos en el Master o en el Satélite, empleando para ello una única ($STD_{Set A}$) o veinte ($STD_{Set B}$) muestras líquidas no selladas de grasa de cerdo como colectivo de estandarización. Sin embargo, no es posible lograr dicho ajuste espectral entre instrumentos a partir de colectivos de estandarización de muestras de características diferentes como ácido oleico ($STD_{Set C}$) o ácido linoléico ($STD_{Set D}$).

Las Figuras 4.5.1 y 4.5.2 muestran las similitudes y las diferencias entre los espectros obtenidos en el instrumento Master y en el Satélite antes y después de estandarizar con los ficheros de estandarización considerados. Únicamente se han indicado las regiones de los espectros con máximas diferencias para mostrar, de forma clara, el efecto que provoca el proceso de estandarización sobre los valores espectrales NIR.

Ambas figuras confirman la mejora lograda en el ajuste de los valores de $\text{Log}(1/R)$ entre instrumentos mediante el uso de una única muestra ($STD_{Set A}$) y varias muestras de estandarización ($STD_{Set B}$) de grasa de cerdo Ibérico. La representación gráfica de los espectros tras la estandarización con los colectivos C y D confirman las conclusiones alcanzadas a partir de los valores de $RMS(C)$ de la Tabla 4.5.2.

Figura 4.5.1. Comparación de espectros de grasa líquida de cerdo Ibérico obtenidos en el instrumento Master y Satélite antes y después de la estandarización con colectivos A y B.

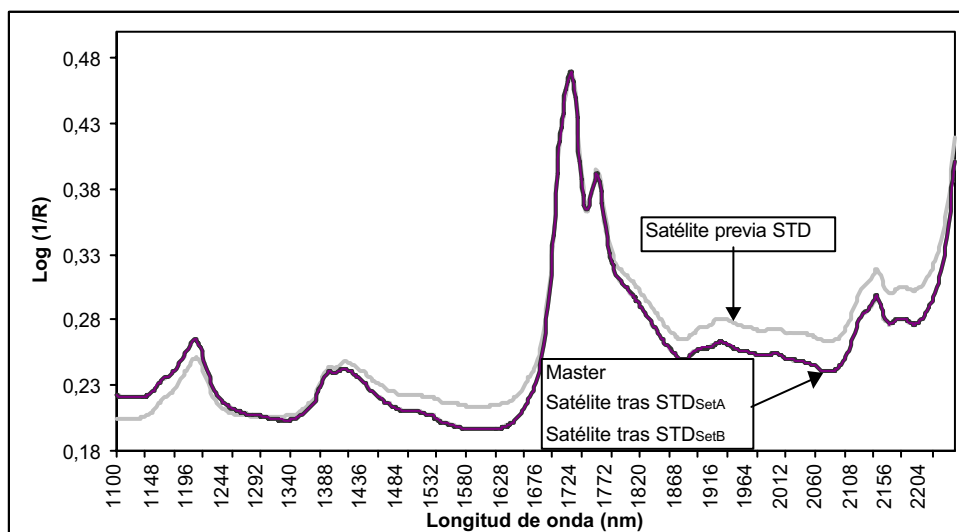
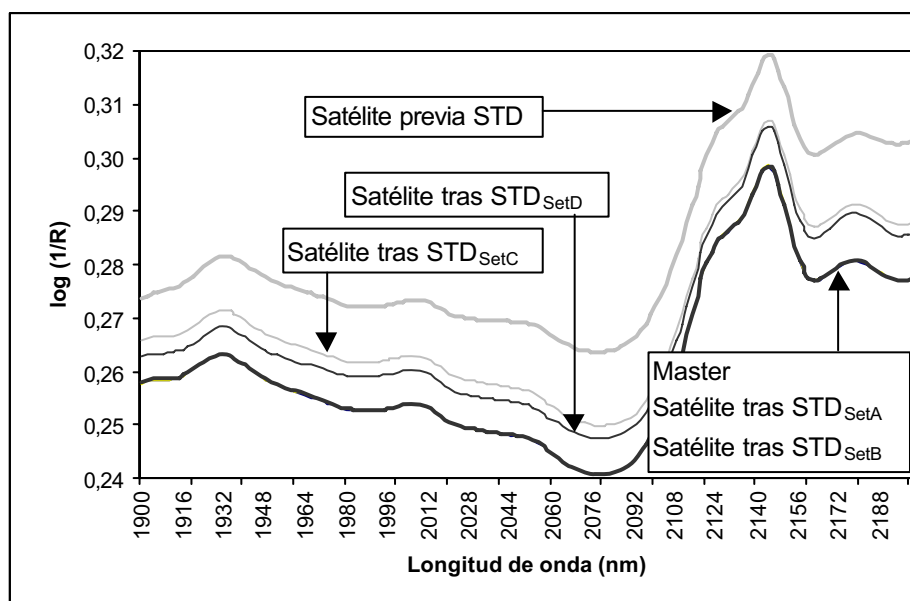


Figura 4.5.2. Comparación de espectros de grasa líquida de cerdo Ibérico obtenidos en el instrumento Master y Satélite antes y después de la estandarización con cuatro colectivos diferentes.



4.5.1.2 Transferencia de ecuaciones

En la Tabla 4.5.3 se indican los estadísticos de las ecuaciones de calibración empleadas en este apartado, desarrolladas a partir de espectros obtenidos en el instrumento Master. Como se puede apreciar, las ecuaciones presentaron una excelente

precisión, con valores de estadísticos de calibración y validación cruzada idénticos a aquellos descritos en el apartado 4.3.1.6 (Tabla 4.3.17).

Tabla 4.5.3. Estadísticos de ecuaciones obtenidas a partir del colectivo de calibración (N=341).

Ácido graso	Media	DT	ETC	R ²	ETVC	r ²
C16:0 (%)	21,00	1,39	0,23	0,97	0,26	0,97
C18:0 (%)	10,68	1,34	0,24	0,97	0,28	0,96
C18:1 (%)	52,24	2,37	0,24	0,99	0,26	0,99
C18:2 (%)	9,40	1,33	0,13	0,99	0,15	0,99

Las Tablas 4.5.4 a 4.5.7 muestran los diferentes estadísticos para las predicciones de los valores de ácidos en las muestras del colectivo de validación E, tanto para el instrumento Master como para el Satélite antes y después de estandarizar con diferentes colectivos (STD_{Set A}; STD_{Set B}; STD_{Set C}; STD_{Set D}).

Tabla 4.5.4. Estadísticos para la predicción de C16:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite, antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	21,28	8,08	21,27	20,90	21,63	22,66
DT (%)	2,02	2,04	20,3	1,70	2,02	2,02
ETD (%)	0,30 ⁽¹⁾	13,20	0,16	0,72	0,39	1,39
Sesgo (%)	-0,16	13,20	0,01	0,37	-0,36	-1,38
r ²	0,99	0,99	0,99	0,91	0,99	0,99
H media	1,67	70,89	1,94	1,73	5,48	11,06

(1) Valor de ETP.

Tabla 4.5.5. Estadísticos para la predicción de C18:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite, antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	10,53	11,23	10,69	10,80	15,01	17,09
DT (%)	1,54	1,54	1,56	1,22	1,54	1,53
ETD (%)	0,29 ⁽¹⁾	0,73	0,28	0,50	4,48	6,56
Sesgo (%)	-0,02	-0,69	-0,16	-0,27	-4,48	-6,55
r ²	0,96	0,98	0,98	0,95	0,98	0,98
H media	1,67	70,89	1,94	1,73	5,48	11,06

(1) Valor de ETP.

Tabla 4.5.6. Estadísticos para la predicción de C18:1 del colectivo de validación E (n=10) en los instrumentos Master y Satélite, antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	52,75	54,02	52,48	52,63	53,10	46,27
DT (%)	3,47	3,32	3,36	2,86	3,34	3,34
ETD (%)	0,35 ⁽¹⁾	1,30	0,39	0,71	0,46	6,49
Sesgo (%)	-0,22	-1,27	0,27	0,12	-0,35	6,49
r ²	0,99	0,99	0,99	0,98	0,99	0,99
H media	1,67	70,89	1,94	1,73	5,48	11,06

(1) Valor de ETP.

Tabla 4.5.7. Estadísticos para la predicción de C18:2 del colectivo de validación E (n=10) en los instrumentos Master y Satélite, antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	9,33	8,50	9,14	9,87	2,99	9,07
DT (%)	1,24	1,11	1,12	0,88	1,10	1,11
ETD (%)	0,11 ⁽¹⁾	0,87	0,30	1,08	6,34	1,28
Sesgo (%)	-0,01	0,83	0,19	-0,55	6,34	1,26
r ²	0,99	0,97	0,97	0,39	0,96	0,97
H media	1,67	70,89	1,94	1,73	5,48	11,06

(1) Valor de ETP.

A pesar de que los resultados obtenidos tras el ajuste espectral confirmaron que únicamente era posible lograr una correcta estandarización entre instrumentos mediante el uso de los colectivos de estandarización A y B, se han indicado los valores obtenidos tras la estandarización con los colectivos C y D para mostrar los resultados desde un punto de vista cuantitativo.

Para todos los estadísticos considerados, los espectros obtenidos en el instrumento Satélite sin estandarizar fueron predichos por las ecuaciones generadas en el Master. Como puede apreciarse en las anteriores tablas, se obtuvieron unos resultados muy diferentes respecto a las predicciones de los espectros tomados en el Master, fundamentalmente a nivel de medias y de distancias *H* media.

Los valores de H medios logrados tras la estandarización con $STD_{Set A}$ (1,94) y $STD_{Set B}$ (1,73) fueron muy similares a los valores de H obtenidos en el instrumento Master (1,67). Sin embargo, la distancia H media de los espectros tras estandarizar con $STD_{Set C}$ (5,48) y $STD_{Set D}$ (11,06) fueron superiores al valor límite recomendado ($H=3$) por Shenk y Westerhaus (1995b).

Los resultados obtenidos mediante el uso de la distancia H media confirmaron aquellos logrados con el estadístico $RMS(C)$ (Tabla 4.5.2), los cuales muestran que únicamente es posible lograr una correcta transferencia de las ecuaciones generadas en el instrumento Master mediante el uso de los colectivos de estandarización A y B.

La estandarización del instrumento Satélite con el colectivo A (estandarización con una única muestra de grasa líquida de cerdo Ibérico) logró mejores valores para todos los estadísticos de predicción de los 4 ácidos grasos considerados (Tablas 4.5.4 a 4.5.7) que la estandarización con el colectivo B (estandarización con 20 muestras de grasa de cerdo Ibérico).

El instrumento Satélite estandarizado con el colectivo A redujo los valores de ETD y $sesgo$ obtenidos antes de estandarizar, para todos los ácidos grasos, logrando idénticos valores para la media y DT que los predichos en el Master. En este caso, los valores de ETD tras estandarizar con $STD_{Set A}$ fueron similares a los $ETVC$ para C16:0 y C18:0 (Tabla 4.5.3) o ligeramente superiores para C18:1 y C18:2 (Tabla 4.5.3).

La estandarización del Satélite con $STD_{Set C}$ (estandarización con una única muestra de ácido C18:1) sólo alcanzó valores aceptables de ETD y $sesgo$ para la predicción de C16:0 y C18:1 (Tablas 4.5.4 y 4.5.6). Sin embargo, mediante el uso de uso del colectivo D para estandarizar (estandarización con una única muestra de ácido C18:2) no se obtuvieron resultados aceptables de los valores de ETD y $sesgo$ para la predicción de ningún ácido graso (Tablas 4.5.4 a 4.5.7). De acuerdo con los resultados de las Tablas 4.5.4 a 4.5.7, se puede concluir que mediante la estandarización con el colectivo A se obtuvieron los mejores resultados de predicción de los 4 ácidos grasos para todos los estadísticos calculados.

De acuerdo con los criterios descritos por Dardenne *et al.* (1992), para considerar que se ha logrado una correcta estandarización y transferencia de ecuaciones

entre instrumentos, los valores de *ETD* entre las predicciones en los instrumentos Master y Satélite han de ser inferiores a los valores de *ETVC* o incluso a *ETC*.

Ha de indicarse que dichas recomendaciones se realizaron para llevar a cabo un proceso de estandarización empleando muestras en polvo, envasadas en cápsulas selladas. Sin embargo, para un proceso de estandarización desarrollado a partir de muestras líquidas no envasadas en cápsulas selladas y de acuerdo con el objetivo final de lograr transferir las ecuaciones generadas a diferentes laboratorios, parece apropiado el uso de otros criterios para evaluar el proceso de clonación y transferencia de ecuaciones entre instrumentos.

Un criterio a emplear podría ser la comparación de los resultados mostrados previamente con aquellos obtenidos en el desarrollo del ensayo colaborativo, descrito en el apartado 4.1.3, para la determinación de las diferencias existentes entre laboratorios al analizar grasa de cerdo Ibérico mediante el método de referencia.

Las diferencias existentes, expresadas como valores de *ETD*, entre un laboratorio de reputada experiencia en la determinación de ácidos grasos en grasas animales y aceites vegetales y otros 5 laboratorios diferentes oscilaron entre 0,17 a 1,56% para C16:0, 0,09 a 1,75% para C18:0, 0,21 a 1,54% para C18:1 y 0,11 a 0,55% para C18:2 con valores de *ETD* medios de 0,84, 0,62, 0,98 y 0,30 para cada ácido graso respectivamente. Como puede observarse en las Tablas 4.5.4 a 4.5.7, los valores de *ETD* entre las predicciones de 2 instrumentos NIRS tras estandarizar con el colectivo A son mucho menores a los valores de *ETD* existentes entre los valores obtenidos por diferentes laboratorios mediante el método de referencia (cromatografía de gases).

Otro criterio empleado para juzgar los resultados obtenidos tras la transferencia de las ecuaciones podría ser la comparación, mediante el estadístico *ETP*, de los valores de referencia frente a los valores predichos, tanto en el instrumento Master como en el Satélite antes y después de desarrollar la estandarización.

La Tabla 4.5.8 muestra los valores predichos para el colectivo de validación E (n=10) en el instrumento Master y Satélite antes y después de estandarizar con el colectivo A (estandarización mediante una única muestra de grasa de cerdo Ibérico).

Tabla 4.5.8. Valores de laboratorio (referencia) y predichos para el colectivo de Validación E en los instrumentos Master y Satélite, antes y después de estandarizar con el colectivo de estandarización A.

C16:0 (%)					C18:0 (%)				
N	Val. Lab.	Val. Pred. Master	Val. Pred. Satélite previa STD	Val. Pred. Satélite tras STD	N	Val. Lab.	Val. Pred. Master	Val. Pred. Satélite previa STD	Val. Pred. Satélite tras STD
1	20,70	21,15	8,20	21,25	1	12,00	11,78	12,28	11,77
2	19,10	19,27	6,31	19,47	2	9,40	9,50	9,81	9,26
3	20,30	20,57	7,24	20,58	3	9,30	9,39	9,85	9,29
4	19,20	19,48	6,06	19,42	4	9,10	9,41	9,98	9,42
5	24,40	23,89	10,60	23,85	5	12,60	13,08	13,71	13,20
6	23,30	23,20	9,74	22,91	6	11,00	11,29	12,20	11,68
7	24,40	24,65	11,62	24,79	7	12,50	12,62	13,40	12,89
8	18,40	18,69	5,32	18,43	8	9,00	8,72	9,83	9,28
9	20,40	20,75	7,63	20,84	9	10,50	10,03	10,89	10,36
10	21,00	21,10	8,09	21,17	10	9,70	9,51	10,31	9,77
ETP		0,30	13,04	0,37	ETP		0,29	0,78	0,35
C18:1 (%)					C18:2 (%)				
N	Val. Lab.	Val. Pred. Master	Val. Pred. Satélite previa STD	Val. Pred. Satélite tras STD	N	Val. Lab.	Val. Pred. Master	Val. Pred. Satélite previa STD	Val. Pred. Satélite tras STD
1	50,50	50,67	52,08	50,53	1	10,30	10,15	9,51	10,15
2	56,40	56,50	57,35	55,85	2	9,30	9,34	8,44	9,09
3	55,40	55,47	56,78	55,25	3	8,70	8,67	7,75	8,39
4	53,20	53,10	54,12	52,56	4	12,20	12,47	11,14	11,83
5	48,00	48,34	49,49	47,89	5	8,20	8,21	7,77	8,40
6	49,70	49,83	51,61	50,05	6	8,60	8,63	7,98	8,61
7	47,80	47,60	49,28	47,69	7	8,40	8,49	7,63	8,26
8	57,00	57,56	58,98	57,50	8	9,30	9,32	8,54	9,18
9	53,10	53,63	54,80	53,27	9	9,50	9,35	8,65	9,30
10	54,20	54,83	55,73	54,22	10	8,70	8,66	7,56	8,18
ETP		0,35	1,53	0,34	ETP		0,11	0,85	0,26

Como era de esperar, las predicciones a partir de espectros sin estandarizar generan valores de *ETP*, para todos los ácidos grasos, muy superiores (13,04% para C16:0; 0,78% para C18:0; 1,53% para C18:1; 0,85% para C18:2) a los valores de *ETP* obtenidos en el instrumento Master (0,30% para C16:0; 0,29% para C18:0; 0,35% para C18:1; 0,11% para C18:2). Sin embargo, a partir de las predicciones de los ficheros del instrumento Satélite estandarizados mediante el colectivo A, se logran valores de *ETP* similares e incluso inferiores (para el caso de C18:1) a los obtenidos en el instrumento Master.

Por tanto, los resultados logrados en este estudio demuestran que es posible transferir ecuaciones NIRS para predecir la composición de ácidos grasos en grasa de

cerdo Ibérico de un instrumento a otro instrumento del mismo fabricante, con una repetibilidad y reproducibilidad similar e incluso superior a aquellas obtenidas mediante el uso del método de referencia. Además, es posible estandarizar instrumentos NIRS mediante el uso de una única muestra de grasa de cerdo Ibérico sin necesidad de mantener dicha muestra envasada en una cápsula sellada.

4.5.2 INSTRUMENTOS NIRS UBICADOS EN LABORATORIOS DIFERENTES

La metodología descrita en este apartado fue la misma que la empleada en el apartado anterior. La única diferencia consistió en que los instrumentos NIRS se encontraban ubicados en laboratorios diferentes por lo que fue necesario un reencapsulado en la toma de espectros de las muestras de grasa. Ello, a priori, supone la incorporación de una nueva fuente de error en todo el proceso de estandarización y transferencia de ecuaciones.

En este caso, como ya se indicó en el apartado 3.5, fueron dos los instrumentos Satélite considerados, localizándose uno de ellos en Zaragoza y otro en Cabrils (Barcelona). A pesar de que tanto los instrumentos Satélite como el Master eran del mismo fabricante, el instrumento Satélite ubicado en Zaragoza estaba equipado, al igual que el instrumento Master, con un módulo de giro mientras que el instrumento de Cabrils disponía de un módulo de transporte.

En ambos instrumentos Satélite se realizó el proceso de clonación con el instrumento Master, empleando los diferentes colectivos de estandarización y validación descritos en el apartado 3.5.2. A continuación se procedió a la transferencia de las ecuaciones descritas en el apartado 3.5.1.

Las Tablas 4.5.9 a 4.5.12 muestran los diferentes estadísticos para las predicciones de los valores de ácidos en las muestras del colectivo de validación E, tanto para el instrumento Master como para el Satélite de Zaragoza antes y después de estandarizar con diferentes colectivos (STD_{Set A}; STD_{Set B}; STD_{Set C}; STD_{Set D}).

Tabla 4.5.9. Estadísticos para la predicción de C16:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Zaragoza), antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	21,28	33,35	20,99	20,70	15,56	18,07
DT (%)	2,02	1,87	1,73	0,53	1,74	1,74
ETD (%)	0,30 ⁽¹⁾	12,28	0,42	1,97	5,72	3,22
Sesgo (%)	-0,16	-12,27	0,28	0,58	5,71	3,20
r ²	0,99	0,99	0,99	0,04	0,99	0,99
H media	1,67	185,38	2,69	1,76	13,01	25,21

(1) Valor de ETP.

Tabla 4.5.10. Estadísticos para la predicción de C18:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Zaragoza), antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	10,53	7,92	10,25	10,04	23,39	23,85
DT (%)	1,54	1,49	1,21	0,31	1,04	1,05
ETD (%)	0,29 ⁽¹⁾	2,63	0,48	1,45	12,87	13,33
Sesgo (%)	-0,02	2,62	0,28	0,49	-12,86	-13,31
r ²	0,96	0,97	0,97	0,16	0,92	0,92
H media	1,67	185,38	2,69	1,76	13,01	25,21

(1) Valor de ETP.

Tabla 4.5.11. Estadísticos para la predicción de C18:1 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Zaragoza), antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	52,75	50,78	52,61	52,79	55,72	39,74
DT (%)	3,47	3,48	2,92	0,33	2,86	2,72
ETD (%)	0,35 ⁽¹⁾	1,98	0,58	3,60	3,03	13,03
Sesgo (%)	-0,22	1,97	0,14	-0,04	-2,96	13,01
r ²	0,99	0,99	0,99	0,95	0,99	0,99
H media	1,67	185,38	2,69	1,76	13,01	25,21

(1) Valor de ETP.

Tabla 4.5.12. Estadísticos para la predicción de C18:2 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Zaragoza), antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	9,33	29,68	8,72	9,65	5,57	20,08
DT (%)	1,24	1,53	1,02	0,27	0,95	1,13
ETD (%)	0,11 ⁽¹⁾	20,35	0,75	1,23	3,79	10,76
Sesgo (%)	-0,01	-20,35	0,61	-0,32	3,76	-10,75
r ²	0,99	0,88	0,88	0,01	0,99	0,88
H media	1,67	185,38	2,69	1,76	13,01	25,21

(1) Valor de ETP.

En el caso del instrumento Satélite localizado en el laboratorio de Cabrils, se repitió el proceso de clonación con el instrumento Master, empleando los diferentes colectivos de estandarización y validación descritos en el apartado 3.5.2. Seguidamente se procedió a la transferencia de las ecuaciones descritas en el apartado 3.5.1.

Las Tablas 4.5.13 a 4.5.16 muestran los diferentes estadísticos para las predicciones del contenido en ácidos grasos mayoritarios de las muestras del colectivo de validación E, tanto para el instrumento Master como para el equipo Satélite de Cabrils, antes y después de estandarizar con los colectivos de estandarización considerados.

Tabla 4.5.13. Estadísticos para la predicción de C16:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Cabrils), antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	21,28	9,77	21,26	21,25	21,81	25,52
DT (%)	2,02	1,96	1,98	0,63	1,91	1,96
ETD (%)	0,30 ⁽¹⁾	11,51	0,46	2,13	0,71	4,27
Sesgo (%)	-0,16	11,50	0,02	0,03	-0,54	-4,25
r ²	0,99	0,95	0,94	0,05	0,94	0,94
H media	1,67	46,11	1,91	1,76	6,39	12,06

(1) Valor de ETP.

Tabla 4.5.14. Estadísticos para la predicción de C18:0 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Cabrils), antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	10,53	10,41	10,49	9,86	17,63	19,54
DT (%)	1,54	1,54	1,55	0,45	1,49	1,52
ETD (%)	0,29 ⁽¹⁾	0,24	0,21	1,50	7,11	9,01
Sesgo (%)	-0,02	0,13	0,04	0,67	-7,10	-9,01
r ²	0,96	0,98	0,98	0,17	0,97	0,97
H media	1,67	46,11	1,91	1,76	6,39	12,06

(1) Valor de ETP.

Tabla 4.5.15. Estadísticos para la predicción de C18:1 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Cabrils), antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	52,75	54,46	52,53	53,76	54,16	42,86
DT (%)	3,47	3,21	3,22	0,53	3,10	3,17
ETD (%)	0,35 ⁽¹⁾	1,73	0,37	3,26	1,45	9,90
Sesgo (%)	-0,22	-1,70	0,23	-1,01	-1,40	9,89
r ²	0,99	0,99	0,99	0,21	0,99	0,99
H media	1,67	46,11	1,91	1,76	6,39	12,06

(1) Valor de ETP.

Tabla 4.5.16. Estadísticos para la predicción de C18:2 del colectivo de validación E (n=10) en los instrumentos Master y Satélite (Cabrils), antes y después de estandarizar con 4 colectivos de estandarización diferentes.

	Master	Satélite previa STD	Satélite tras STD _{Set A}	Satélite tras STD _{Set B}	Satélite tras STD _{Set C}	Satélite tras STD _{Set D}
Media (%)	9,33	9,76	9,18	2,42	8,17	9,67
DT (%)	1,24	1,22	1,23	1,22	1,22	0,47
ETD (%)	0,11 ⁽¹⁾	0,50	0,29	6,91	1,19	1,07
Sesgo (%)	-0,01	-0,43	0,15	6,91	1,16	-0,35
r ²	0,99	0,96	0,96	0,94	0,95	0,29
H media	1,67	46,11	1,91	6,39	12,06	1,76

(1) Valor de ETP.

Al igual que en la estandarización y transferencia de ecuaciones entre instrumentos ubicados en el mismo laboratorio, en la estandarización del instrumento

Satélite de Zaragoza y de Cabrils se obtuvo unos mejores estadísticos para todos los ácidos grasos considerados con el colectivo de estandarización A.

Una prueba del éxito en la estandarización con dicho colectivo se observa en el valor de H , el cual presentó valores muy altos en los instrumentos Satélite antes de la clonación (H Zaragoza = 185,38; H Cabrils = 46,11), disminuyendo hasta valores inferiores a 3 tras estandarizar con el colectivo A (H Zaragoza = 2,69; H Cabrils = 1,91). Además, al emplear dicho colectivo de estandarización, los valores de ETD obtenidos para los diferentes ácidos grasos en ambos instrumentos Satélites disminuyeron sensiblemente tras la clonación, llegando a ser incluso del mismo orden a los valores de ETP en el instrumento Master.

Esto no se observó en el caso del ácido graso C18:2 donde en ambos instrumentos Satélite y en particular, en el ubicado en Zaragoza, los valores de ETD fueron superiores al doble del valor de ETP . Sin embargo, dado que el valor de ETD para este ácido graso en dicho instrumento antes de estandarizar fue muy alto (ETD C18:2 = 20,35%), se puede considerar como adecuada la precisión obtenida tras la clonación.

Observando con más detalle las tablas 4.5.9 a 4.5.16, se puede apreciar que en la mayoría de los casos, los valores de ETD de los diferentes ácidos graso en los instrumentos Satélite antes de estandarizar son similares a los valores de *sesgo*. Ello muestra que la mayor parte de las diferencias existentes entre el instrumento Master y los instrumentos Satélites son diferencias sistemáticas controladas como son, fundamentalmente, las existentes entre los instrumentos NIRS.

4.5.3 REPETIBILIDAD Y REPRODUCIBILIDAD DE LA PREDICCIÓN DE ÁCIDOS GRASOS MEDIANTE NIRS

Como resultado de la estandarización descrita en los apartados anteriores, se obtuvo la composición en ácidos grasos mayoritarios de las mismas muestras de grasa (colectivo de validación E), aportada por 4 instrumentos NIRS diferentes ubicados en 3 laboratorios (en Córdoba, en Zaragoza y en Cabrils). Dado que los espectros de cada muestra se obtuvieron por duplicado, se obtuvieron las predicciones de cada uno de dichos espectros considerándolos como repeticiones del análisis de una misma muestra.

Esto permitía abordar un estudio de repetibilidad y reproducibilidad de determinación de ácidos grasos mediante NIRS.

En la Tabla 4.5.17 se incluyen los valores medios de las predicciones de cada muestra obtenidas en los diferentes instrumentos junto con el valor obtenido mediante el método de referencia. Aclarar que los valores predichos en el laboratorio A1 (Lab A1) corresponderían a los obtenidos en el instrumento Master, los predichos en el laboratorio A2 (Lab A2) a los obtenidos en el equipo Satélite instalado junto al instrumento Master en el laboratorio de Córdoba, los predichos en el laboratorio B (Lab B) a los obtenidos en el instrumento Satélite de Zaragoza y los predichos en el laboratorio C (Lab C) a los generados en el equipo Satélite de Cabrils.

Tabla 4.5.17. Valores de laboratorio (referencia) y predichos para el colectivo de Validación E en los diferentes instrumentos (Master y Satélites).

C16:0 (%)						C18:0 (%)					
N	Val Ref.	Pred. Lab A1	Pred. Lab A2	Pred. Lab B	Pred. Lab C	N	Val Ref.	Pred. Lab A1	Pred. Lab A2	Pred. Lab B	Pred. Lab C
1	20,70	21,15	21,25	20,85	20,37	1	12,00	11,78	11,77	11,29	11,93
2	19,10	19,27	19,47	19,31	18,90	2	9,40	9,50	9,26	9,63	9,01
3	20,30	20,57	20,58	20,35	20,61	3	9,30	9,39	9,29	8,89	9,58
4	19,20	19,48	19,42	19,43	19,07	4	9,10	9,41	9,42	9,30	9,20
5	24,40	23,89	23,85	23,50	23,62	5	12,60	13,08	13,20	12,03	12,99
6	23,30	23,20	22,91	22,39	23,19	6	11,00	11,29	11,68	10,77	11,31
7	24,40	24,65	24,79	23,84	24,50	7	12,50	12,62	12,89	12,13	12,56
8	18,40	18,69	18,43	18,80	19,33	8	9,00	8,72	9,28	9,10	8,99
9	20,40	20,75	20,84	20,68	21,42	9	10,50	10,03	10,36	9,94	9,95
10	21,00	21,10	21,17	20,76	21,56	10	9,70	9,51	9,77	9,41	9,43
ETP		0,30	0,37	0,49	0,56	ETP		0,29	0,35	0,41	0,30
C18:1 (%)						C18:2 (%)					
N	Val Ref.	Pred. Lab A1	Pred. Lab A2	Pred. Lab B	Pred. Lab C	N	Val Ref.	Pred. Lab A1	Pred. Lab A2	Pred. Lab B	Pred. Lab C
1	50,50	50,67	50,53	51,09	50,54	1	10,30	10,15	10,15	9,37	10,39
2	56,40	56,50	55,85	55,79	56,05	2	9,30	9,34	9,09	8,54	9,36
3	55,40	55,47	55,25	54,95	55,17	3	8,70	8,67	8,39	7,79	8,58
4	53,20	53,10	52,56	53,08	53,12	4	12,20	12,47	11,83	11,03	12,17
5	48,00	48,34	47,89	48,51	48,33	5	8,20	8,21	8,40	7,64	8,28
6	49,70	49,83	50,05	50,25	49,63	6	8,60	8,63	8,61	8,07	8,21
7	47,80	47,60	47,69	48,37	47,98	7	8,40	8,49	8,26	8,67	8,40
8	57,00	57,56	57,50	56,47	56,92	8	9,30	9,32	9,18	8,91	8,71
9	53,10	53,63	53,27	53,20	53,18	9	9,50	9,35	9,30	9,33	9,00
10	54,20	54,83	54,22	54,39	54,36	10	8,70	8,66	8,18	7,88	8,67
ETP		0,35	0,34	0,46	0,19	ETP		0,11	0,26	0,72	0,28

En dicha tabla también se incluyen los valores de *ETP* para cada una de las predicciones realizadas en el instrumento Master y en los diferentes instrumentos Satélite. Como puede apreciarse, estos resultados de *ETP* son similares a los obtenidos de *ETD* en las 4.5.4 a 4.5.16.

Además, estos valores de *ETP* obtenidos en los diferentes laboratorios NIRS (instrumentos) son del mismo orden para cada uno de los ácidos grasos, a excepción de los valores obtenidos para el ácido C18:2. En este caso, este estadístico fue excesivamente elevado para las predicciones del laboratorio B (instrumento Satélite de Zaragoza).

A partir de los datos por duplicado (repeticiones) de cada una de las muestras en los diferentes laboratorios se determinó, mediante la metodología descrita en el apartado 3.1.3.2, los estadísticos acerca de la repetibilidad y reproducibilidad existente en el análisis de ácidos grasos en grasa de cerdo Ibérico mediante la tecnología NIRS. Es decir, se consideraron los resultados como si hubieran sido obtenidos dentro de un ensayo colaborativo como el descrito para la determinación de estos mismos analitos mediante cromatografía de gases.

Tras realizar los test numéricos, únicamente se detectó un dato anómalo mediante el test de Cochran (valor de C18:1 de la muestra 2 para el laboratorio A2). Al realizar el test de Grubbs sencillo y doble ningún dato fue detectado como anómalo.

Tras eliminar los datos provenientes de dicha muestra, se calcularon los diferentes estimadores de la repetibilidad y la reproducibilidad. En la Tabla 4.5.18, se incluyen los valores de desviación típica de la repetibilidad y reproducibilidad, los estadísticos *repetibilidad* y *reproducibilidad* junto con los datos de los estimadores RSD_r y RSD_R .

Tabla 4.5.18. Parámetros estadísticos de repetibilidad y reproducibilidad

Ácido graso	Media Global y	DT Repetibilidad s_r	DT Reproducibilidad s_R	Repetibilidad r	Reproducibilidad R	RSD_r (%)	RSD_R (%)
C16:0 (%)	21,20	0,24	0,35	0,68	0,99	1,15	1,67
C18:0 (%)	10,49	0,20	0,33	0,55	0,91	1,95	3,08
C18:1 (%)	52,60	0,23	0,35	0,66	0,99	0,45	0,67
C18:2 (%)	9,09	0,25	0,40	0,70	1,12	2,79	4,38

Como ya se indicó, el estadístico de la desviación típica de la repetibilidad (s_r) es un estimador del error de cada uno de los laboratorios (error intralaboratorio). Los valores de dicho estadístico para cada uno de los ácidos grasos fue similar a los valores de *ETL* estimados sobre duplicados de 20 muestras y de *ETVC* de las diferentes ecuaciones NIRS.

Estos valores de s_r fueron, sin embargo, superiores a los obtenidos en el ensayo colaborativo realizado mediante cromatografía de gases. Ello podría ser debido a la ya comentada minimización del error intralaborial realizada en los laboratorios participantes en el estudio colaborativo de la técnica cromatográfica, aspecto éste que no se controló de igual medida en el laboratorio responsable de aportar los datos de referencia.

De acuerdo con los valores del estadístico repetibilidad (r) las máximas diferencias absolutas entre 2 resultados obtenidos en el mismo instrumento NIRS variarían entre 0,55 y 0,70% para los distintos ácidos grasos mayoritarios.

Sin embargo, a diferencia de los resultados obtenidos en el ensayo colaborativo realizado mediante cromatografía de gases, los valores de desviación típica de la reproducibilidad (s_R), estimador del error interlaboratorio, presentaron valores ligeramente superiores a los de desviación típica de repetibilidad. Es de interés recordar que, como fue descrito en la Tabla 4.1.14, los valores de s_R en la determinación de ácidos grasos mediante cromatografía de gases fueron del orden de 4 a 8 veces superiores a los estimadores de s_r . Como se indica en la Tabla 4.5.18, al realizar estas determinaciones mediante cromatografía de gases, las diferencias entre dichos estadísticos para un mismo ácido graso fueron como máximo de orden 2.

De acuerdo con los valores de reproducibilidad (R) incluidos en la Tabla 4.5.18, las máximas diferencias absolutas entre 2 resultados obtenidos en diferentes instrumentos NIRS serían, aproximadamente del 1% para los distintos ácidos grasos mayoritarios. Estas diferencias podrían ser de hasta 2,5% al analizar una misma muestra en dos laboratorios diferentes mediante cromatografía de gases.

Como ya se indicó en el apartado 4.1.3, la norma 963.22 de la AOAC (1999) especifica la precisión que ha de tener el método de análisis empleado en este estudio, indicando los valores máximos de repetibilidad y reproducibilidad que han de

presentar los ácidos grasos mayoritarios (>5%) en las muestras de grasas o aceites. Así, los valores de r y R han de ser inferiores a 1 y 3 respectivamente, mientras que los valores de RSD_r no deben ser superiores a 3% y los valores de RSD_R no deben superar el 10%.

Como se observa en la Tabla 4.5.18, a pesar de que la repetibilidad fue superior a la obtenida mediante cromatografía de gases, los valores de r y RSD_r para los 4 ácidos grasos analizados en este estudio son inferiores a los valores recomendados como máximos por la AOAC. Además, los valores de R y RSD_R para dichos ácidos fueron también inferiores a los máximos recomendados por la AOAC e incluso inferiores a los obtenidos mediante cromatografía de gases, a excepción del ácido C18:2.

Por todo ello se puede concluir que la determinación del contenido en ácidos grasos en grasa de cerdo Ibérico presentar una precisión intralaboratorial igual a la del laboratorio que aportó los datos de referencia mediante cromatografía de gases. Sin embargo, la precisión interlaboratorial obtenida entre instrumentos NIRS fue, en general, muy superior a la existente entre laboratorios que empleen como técnica de análisis la cromatografía de gases para determinar el contenido de ácidos grasos mayoritarios en grasa de cerdo Ibérico.

5 CONCLUSIONES

1. Las diferentes zonas anatómicas de toma de muestra de grasa de la canal estudiadas presentan diferencias en la composición en ácidos grasos C16:0, C18:0, C18:1 y C18:2.
2. Para la clasificación de canales de cerdo Ibérico en función del régimen alimenticio, empleando como criterio la composición de los ácidos grasos mayoritarios, se recomienda obtener la muestra de tejido adiposo subcutáneo de la zona de coxal, dada la menor variabilidad en composición de ácidos grasos entre animales de un mismo lote y dada su mayor facilidad de toma de muestras en la canal.
3. Es deseable controlar de forma individual la canal de cada animal, ya que se han detectado diferencias importantes en cuanto a composición en ácidos grasos mayoritarios entre animales con un mismo régimen alimenticio durante la etapa de cebo. En el caso de que se opte por obtener una única muestra de grasa fundida representativa de un determinado lote se ha de muestrear, al menos, el 50% de los animales de dicho lote.
4. El ensayo colaborativo realizado para la evaluación de errores intra e interlaboratoriales en la determinación de ácidos grasos mediante cromatografía de gases en grasa de cerdo Ibérico pone de manifiesto que los valores de repetibilidad y reproducibilidad obtenidos son inferiores a los máximos recomendados en la bibliografía para dicho análisis, lo cual muestra la elevada precisión de los laboratorios participantes.
5. Para asegurar una alta repetibilidad de los datos espectroscópicos NIRS de grasa fundida se recomienda que la temperatura de las muestras durante su preparación y análisis se mantenga constante a 40°C, realizar el análisis por duplicado considerando un valor límite de RMS de 4000 $\mu\log(1/R)$ y evitar la formación de burbujas en el interior de la cápsula.
6. La repetibilidad del análisis NIRS mediante sonda de fibra óptica en tejido adiposo es inferior a la obtenida en grasa fundida. Para minimizar el error del dato espectroscópico obtenido en esta forma de presentación de muestra se recomienda el mantener la sonda en posición vertical para lograr un ángulo de 90° entre la dirección de la radiación incidente y la superficie de la muestra, obtener un mínimo

de 3 espectros en tres posiciones diferentes de una misma muestra y considerar un valor límite de RMS de $45.000 \mu\log(1/R)$.

7. La conservación de las muestras de grasa fundida de cerdo Ibérico bajo congelación provoca una clara influencia sobre la región Visible (400 a 1100 nm) de la medida espectral realizada. Este factor no provoca alteraciones de los datos espectroscópicos de la región NIR por lo que se recomienda emplear únicamente dicha región para el desarrollo de ecuaciones de calibración o modelos de clasificación.
8. Las ecuaciones NIRS obtenidas para la predicción de los cuatro ácidos grasos mayoritarios en grasa fundida de cerdo Ibérico, obtenidas mediante la modalidad de doble transmisión, poseen una exactitud y precisión inmejorables, siendo semejantes a las del método de referencia (cromatografía de gases). Además presentan una gran robustez al predecir con exactitud y precisión colectivos analizados en periodos diferentes al del desarrollo de las calibraciones.
9. Estas ecuaciones poseen un elevado valor científico y económico, ya que han sido obtenidas utilizando un elevado número de muestras procedentes de controles de campo de dos años consecutivos, y por son representativas de diferentes regímenes alimenticios, genética, condiciones agroclimáticas diferentes, etc.
10. Las ecuaciones obtenidas para la determinación de ácidos grasos mayoritarios en tejido adiposo mediante el uso de una sonda de fibra óptica presentan una alta precisión y exactitud si bien ambos estadísticos serían mejorables. Para disminuir el error en dichas ecuaciones se recomienda el incrementar la repetibilidad del dato espectroscópico NIR, incrementar el número de muestras del colectivo de calibración y asimismo realizar el análisis por cromatografía de gases de la misma muestra de grasa donde se haya posicionado la sonda de fibra óptica.
11. El método de regresión considerado presenta escasa influencia sobre el error de predicción de las ecuaciones desarrolladas. Así, es posible obtener calibraciones de ácidos grasos mayoritarios tanto en grasa fundida como en tejido adiposo mediante regresión lineal múltiple, utilizando un máximo de 10 longitudes de onda, con similar exactitud y precisión que aquellas desarrolladas mediante regresión por

mínimos cuadrados parciales. Este aspecto es de gran interés para el diseño de equipos miniaturizados o instrumentos más económicos.

12. El tratamiento de derivadas considerado durante el desarrollo de las ecuaciones para determinar el contenido en ácidos grasos mayoritarios presenta una notable influencia sobre el error de predicción de dichas ecuaciones. En la presente Tesis Doctoral, las mejores ecuaciones para la predicción del contenido en ácidos grasos en grasa fundida se obtuvieron con los tratamientos de derivadas (1,10,10,1) para el ácido C16:0, (2,10,5,1) para el ácido C18:0, (2,5,5,1) para el ácido C18:1 y (3,10,10,1) para el ácido C18:2.
13. En la aplicación del algoritmo CENTER, se recomienda aplicar un tratamiento de derivación (2,10,5,1) sobre los datos espectrales NIR y utilizar un valor de H inferior a 3 para la detección de espectros anómalos, siempre y cuando se tengan controladas las diferentes fuentes de error instrumentales y ambientales.
14. Para evitar errores inesperados durante el análisis NIRS de rutina, se recomienda el análisis de una muestra de grasa control de composición conocida, previamente al análisis del resto de muestras. Los espectros de dicha muestra de grasa son de gran utilidad para modelizar gran parte de las variaciones instrumentales, operaciones y/o ambientales que afectan al análisis NIRS de grasa fundida. Entre dichas fuentes de variación destacan los cambios en la precisión de la longitud de onda del instrumento NIR, factor que influye sobre los datos espectroscópicos provocando variaciones inesperadas sobre el valor del estadístico H .
15. Es posible la determinación del contenido en ácidos grasos minoritarios en grasa fundida de cerdo Ibérico con una alta exactitud y precisión, pudiendo lograr una mejora en dichos términos mediante la disminución del método de referencia y el incremento de muestras del colectivo de calibración. Estos ácidos grasos minoritarios también se pueden determinar en tejido adiposo mediante un instrumento NIRS dotado de una sonda de fibra óptica si bien, los errores de predicción son mayores a los obtenidos en grasa fundida.
16. Los modelos cualitativos generados para la clasificación de canales de cerdo Ibérico en función de la alimentación, a partir de datos espectrales NIR *per se* presentaron

en general menor error de clasificación que los desarrollados a partir de su composición en ácidos grasos.

17. Los modelos cualitativos para la evaluación del régimen alimenticio desarrollados a partir de muestras de grasa fundida de cerdos Ibéricos, presentaron menor error que aquellos obtenidos a partir de muestras de tejido adiposo, independientemente del método de clasificación multivariante considerado.
18. Se recomienda el uso del método de análisis discriminante basado en regresión PLS2 por haber mostrado menores errores de clasificación tanto para grasa fundida como para tejido adiposo y por la oportunidad que presenta para su aplicación en rutina, al estar incorporado en el mismo programa utilizado para la recogida de datos espectrales.
19. El algoritmo de estandarización de instrumentos NIRS descrito por Shenk y Westerhaus se ha mostrado muy útil para anular las diferencias ópticas existentes entre 4 equipos ubicados tanto en el mismo laboratorio como en 3 laboratorios diferentes, empleando para ello una única muestra no sellada de grasa fundida de cerdo Ibérico.
20. Las calibraciones NIRS desarrolladas para la determinación del contenido en ácidos grasos mayoritarios en grasa fundida han sido transferidas con éxito desde el instrumento Master en el que se generaron a 3 instrumentos Satélites ubicados 2 de ellos en laboratorios diferentes al del Master.
21. La elevada repetibilidad y reproducibilidad obtenida en el ensayo colaborativo de determinación del contenido en ácidos grasos mayoritarios en grasa fundida de cerdo Ibérico mediante las ecuaciones NIRS desarrolladas, en comparación con los valores de repetibilidad y reproducibilidad del método de referencia, demuestran la capacidad de la tecnología NIRS para sustituir a la cromatografía de gases como método oficial de clasificación de canales de cerdo Ibérico en función de la alimentación.

6 BIBLIOGRAFÍA

- Abrams, S.M. 1985. Populations. En Near Infrared Reflectance Spectroscopy (NIRS): Analysis of forage quality. Marten, G.C., Shenk, J. and Barton, F.E. (eds.). USDA. ARS. Agriculture Handbook N° 643. Washington, USA. p. 37.
- AOAC. 1999a. Official Methods of Analysis of the Association of Analytical Chemists. Method 963.22. Methyl esters of fatty acids in oils and fats. AOAC International, Gaithersburg, Maryland, USA.
- AOAC. 1999b. Official Methods of Analysis of the Association of Analytical Chemists. Appendix D: Guidelines for collaborative study procedures to validate characteristics of a method of analysis. AOAC International, Gaithersburg, Maryland, USA.
- AOCS. 1998. Official methods and recommended practices of the American Oil Chemists' Society. 5th Edition. AOCS Press, Champaign, Illinois, USA.
- ASTM. 1990. Standard definitions of terms and symbols relating to molecular spectroscopy. American Society for Testing and Materials vol. 14.01, Standard E131-90. ASTM, West Conshohocken, PA, USA.
- ASTM. 1994. Standard practices for infrared, multivariate, quantitative analysis. ASTM Annual book of standards. Standard E1655-94. ASTM, West Conshohocken, PA, USA.
- ASTM. 2000. Standard practices for infrared, multivariate, quantitative analysis. ASTM Annual book of standards. Standard E1655-00. ASTM, West Conshohocken, PA, USA.
- Baena, F. Comunicación personal. Dpto. de Producción Animal (E.T.S.I.A.M.). Universidad de Córdoba.
- Barnes, R.J., Dhanoa, M.S. and Lister, S.J. 1989. Standard Normal Variate transformation and De-Trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.*, 43:772-777.
- Barnes, R.J., Dhanoa, M.S. and Lister, S.J. 1994. Correction to the description of Standard Normal Variate (SNV) and De-Tren (DT) transformations in practical spectroscopy with applications in food and beverage analysis. *NIR News*, 5, 3:6.
- Barnett, V. and Lewis, T. 1994. Outliers in statistical data. John Wiley and Sons, New York, NY, USA.
- Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. 1995. Leaping ahead with near infrared spectroscopy. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia.
- Ben-Gera, I. and Norris, K. 1968a. Determination of moisture content in soybeans by direct spectrophotometry. *Isr. J. Agr. Res.*, 18:125.
- Ben-Gera, I. and Norris, K. 1968b. Direct spectrophotometric determination of fat and moisture in meat products. *J. Food Sci.*, 33:64.
- Benito J., Vázquez C., Menaya C., Ferrera J.L. y García Casco J., 1998. Clasificación de calidades en función del perfil de ácidos grasos: Ejercicio comparativo en 4 campañas con cerdos Ibéricos cebados en montanera pura. *Solo Cerdo Ibérico*, 1:93-101.

- Bertran, E., Blanco, M., Coello, J., Iturriaga, H., MasPOCH, S. and Montoliu, I. 2000. Near infrared spectrometry and pattern recognition as screening methods for the authentication of virgin olive oils of very close geographical origins. *J. Near Infrared Spectrosc.*, 8:45-52.
- Bertrand, D. 1993. Data pre-treatment and original analysis in spectroscopy. En *Advanced Comet Chemometrics School*. Libramont Belgium April 26-28 1993.
- Bertrand, D. and Dufour, E. 2000. *La spectroscopie infrarouge et ses applications analytiques*. Editions TEC & DOC, Paris, France.
- Bewig, K.M., Clarke, A.D., Roberts, C. and Unklesbay, N. 1994. Discriminant analysis of vegetable oils by near-infrared reflectance spectroscopy. *J. Am. Oil Chem. Soc.*, 71:195-200.
- Bewig, K.M., Clarke, A.D., Roberts, C. and Unklesbay, N. 1995. Discriminant analysis of vegetable oils using near infrared spectroscopy. En *Leaping ahead with near infrared spectroscopy*. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia. p. 324-328.
- Biston, R. and Bartiaux-Thill, N. 1991. Proceedings of the 3rd International Conference on Near Infrared Spectroscopy. Biston, R. and Bartiaux-Thill, N. (eds.). Agricultural Research Centre, Gembloux, Belgium.
- Blanco, M., Coello, J., Iturriaga, H., MasPOCH, S. and Gonzalez, R. 1998. *Mikrochim. Acta*, 128:235.
- Blank, T.B., Sum, S.T. and Brown, S.D. 1996. Transfer of near infrared multivariate calibrations without standards. *Anal. Chem.*, 68: 2987-2995.
- Bockisch, M. 1998. Composition, structure, physical data and chemical reactions of fats and oils, their derivatives and their associates. En *Fats and oils Handbook*. AOCS Press, Champaign, Illinois, USA. p. 53-120.
- BOE. 1988. Orden de 7 de noviembre de 1998 por la que se homologa el contrato-tipo de compraventa de cerdos Ibéricos y sus cruces, con destino a su sacrificio y elaboración, para la campaña 1998-89. Boletín Oficial del Estado nº 268 de 8 de noviembre de 1988.
- BOE. 1998. Orden de 22 de julio de 1998 por la que se homologa el contrato-tipo de compraventa de cerdos Ibéricos cebados con destino a sacrificio y elaboración, que regirá hasta el 31 de agosto de 1998. Boletín Oficial del Estado nº 181 de 30 de julio de 1998.
- BOE. 2000. Orden de 19 de octubre de 2000 por la que se homologa el contrato-tipo de compraventa de cerdos Ibéricos cebados con destino a sacrificio y elaboración, que regirá hasta el 31 de agosto de 2001. Boletín Oficial del Estado de 258 de 27 de octubre de 2000.
- Bouveresse, E. and Massart, D.L. 1996. Standardisation of near-infrared spectrometric instruments: A review. *Vibrational Spectroscopy*, 11:3-15.
- Burns, D.A. and Ciurczak, E.W. 1992. *Handbook of near infrared analysis*. Burns, D.A. and Ciurczak, E.W. (eds.). Marcel Dekker, New York, NY, USA.
- Buxadé, C y Daza, A. 2000. *Porcino Ibérico: aspectos claves*. Ediciones Mundi-Prensa, Madrid, España.

- Carrapiso, A.I., Ventanas, J., Jurado, A. and García, C. 2001. An electronic nose to classify Iberian pig fats with different fatty acid composition. *J. Am. Oil. Chem. Soc.*, 78:415-418.
- Casillas, M. 1995. Metodologías de caracterización de grasa de cerdo Ibérico para el control de calidad de sus productos. Tesis Doctoral. Universidad de Córdoba.
- Cava, R., Ruiz, J., López-Bote, C., Martín, L., García, C., Ventanas, J. and Antequera, T. 1997. Influence of finishing diet on fatty acid profiles of intramuscular lipids, triglycerides and phospholipids in muscles of the Iberian pig. *Meat Sci.*, 45:263-270.
- Cava, R., Ruiz, J., Ventanas, J. and Antequera, T. 1999. Oxidative and lipolytic changes during ripening of Iberian hams as affected by feeding regime: extensive feeding and alpha-tocopheryl acetate supplementation. *Meat Sci.*, 52:165-172.
- Cava, R., Ventanas, J., Tejeda, J.F., Ruiz, J. and Antequera, T. 2000. Effect of free range rearing and alpha-tocopherol and copper supplementation on fatty acid profiles and susceptibility to lipid oxidation of fresh meat from Iberian pigs. *Food Chem.*, 68:51-59.
- Cert, A., Moreda, W., León-Camacho, M. y Pérez-Camino, M.C. 1996. Determinación de absorción de luz UV a 232 nm, composición de ácidos grasos, trilinoleína y triglicéridos con número equivalente de carbonos igual a 42, en aceites de oliva y de orujo de oliva: Determinación de precisión de los métodos analíticos mediante el estudio estadístico de los resultados de un análisis colaborativo. *Grasas y Aceites*, 47:401-410.
- Cobo C. 1999. Estudio de viabilidad de aplicación de la tecnología NIRS al control de calidad del aceite de oliva. Trabajo Profesional Fin de Carrera. E.T.S.I.A.M. Universidad de Córdoba.
- Cobo, C., Garrido-Varo, A., Sánchez-Pineda, T., Alcalá, R., Horcas, J.M. y Jiménez, A. 2001. Estudio de viabilidad de aplicación de la tecnología NIRS al control de calidad de aceite de oliva. En *Expoliva* (ed.) Gráficas Francisco del Moral, S.A. Jaén. España. p. 87-96.
- Czarnik-Matusiewicz, H.W. and Korniewicz, A. 1995. The use of InfraAlyzer 500 for the determination of iodine value in pig back fat. En *Leaping ahead with near infrared spectroscopy*. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia. p. 300-302.
- Che Man, Y.B. and Moh, M.H. 1998. Determination of free fatty acids in palm oil by near infrared reflectance spectroscopy. *J. Am. Oil Chem. Soc.*, 75:557-562.
- Chen, Y.S. and Chen, A.O. 1995. Quality analysis and purity examination of edible vegetable oils by near infrared transmittance spectroscopy. En *Leaping ahead with near infrared spectroscopy*. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia. p. 316-323.
- Cho, R.K. and Iwamoto, M. 1989. The purity identification of sesame oil by near infrared reflectance spectroscopy. En *Proceedings of the 2nd International near infrared spectroscopy conference*. Iwamoto, M. and Kawano, S. (eds.). Korian Publishing Co, Japan. p. 142-147.

- Cho, S.Y., Kim, J.Y. and Rhee, C. 1998. Determination of rancidity of soybean oil by near infrared spectroscopy. *J. Near Infrared Spectrosc.*, 6:A349-A354.
- Dardenne, P., Biston, R. and Sinnaeve, G., 1992. Calibration transferability across NIRS instruments. En *Near infra-red spectroscopy. Bridging the Gap between Data Analysis and NIR Applications*. Hildrum, K.I., Isaksson, T, Naes, T. and Tandberg, A. (eds.). Ellis Horwood, Chichester, UK. p. 1-10.
- Dardenne, P. 2000a. Editorial. Authentification des productions agricoles et agro-alimentaires: nouveau challenge analytique. *Biotechnol. Agron. Soc. Environ.*, 4:195.
- Dardenne, P. 2000b. Transferts d'équation d'étalonnage et mise en réseau d'instruments. En *La spectroscopie infrarouge et ses applications analytiques*. Editions TEC & DOC, Paris, France. p. 371-379.
- Davies, A.M.C. 1999. NIR instrumentation companies: the story so far. *NIR News*, 10:14-15.
- Davies, A.M.C. and Williams P. 1996. *Near Infrared Spectroscopy: The Future Waves*. Davies, A.M.C. and Williams P. (eds.) NIR Publications, Chichester, UK.
- Davies, A.M.C. and Giangiacomo, R. 2000. En *Near infrared spectroscopy: Proceedings of the 9th International Conference*. Davies, A.M.C. and Giangiacomo, R. (eds.). NIR Publications, Chichester, UK.
- Davies, A.M.C. and Cho, R.K. 2002. *Near Infrared Spectroscopy: Proceedings of the 10th International Conference*. Davies, A.M.C. and Cho, R.K. (eds.). NIR Publications, Chichester, UK.
- De la Hoz, L., López, M.O., Cambero, M.I., Martín-Álvarez, P.J., Gallardo, E. and Ordoñez, J.A. 1993. Fatty acid of Iberian pig liver as affected by diet. *Archiv fur Lebensmittelhygiene*, 44: 81-104.
- De Maesschalek, R., Jouan-Rimbaud, D. and Massart, D.L. 2000. Tutorial: the Mahalanobis distance. *Chem. Intell. Lab. Syst.*, 50:1-18.
- De Noord, O.E. 1994. Multivariate calibration standardization. *Chem. Intell. Lab. Syst.*, 25:85-97.
- De Pedro, E. 1987. Estudio de los factores sexo y peso de sacrificio sobre las características de la canal de cerdo Ibérico. Tesis Doctoral. Universidad de Córdoba.
- De Pedro, E. 1995. La calidad: parámetros que la definen. En *El cerdo Ibérico. Jornadas Técnicas COVAP*. p. 63-81.
- De Pedro, E. 2000. Calidad de las canales y de los productos del cerdo Ibérico: técnicas de control y criterios de calidad. En *Porcino Ibérico: aspectos claves*. Ediciones Mundi-Prensa, Madrid, España. p. 589-621.
- De Pedro, E. y Secondi, F. 1991. Efecto de la raza y de la alimentación en la composición de la grasa subcutánea del jamón de cerdo Ibérico. *ITEA*, 11:455-457.
- De Pedro, E.J., Garrido, A., Bares, I., Casillas, M. and Murray, I. 1992. Application of near infrared spectroscopy for quality control of Iberian pork industry. En *Near infra-red spectroscopy. Bridging the Gap between Data Analysis and NIR*

- Applications. Hildrum, K.I., Isaksson, T, Naes, T. and Tandberg, A. (eds.). Ellis Horwood, Chichester, UK. p. 345-348.
- De Pedro, E., Garrido A., Lobo, A., Dardenne, P. and Murray, I. 1995. Objective classification of Iberian pig carcasses: GC versus NIR. En *Leaping ahead with near infrared spectroscopy*. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia. p. 291-295.
- De Pedro, E., Casillas, M. and Miranda, C.M. 1996. Microwave oven application in the extraction of fat from the subcutaneous tissue of Iberian pig ham. *Meat Sci.*, 45:45-51.
- De Pedro, E., N. Núñez, A. Garrido, J. García, L. Silió, M.C. Rodríguez and J. Rodríguez. 2001. Qualitative analysis of NIRS spectral data to identify Iberian pig feeding types. En *Proceedings of the 6th Int. symposium of food authenticity and safety*. Eurofins Scientific (ed.). Eurofins Scientific, Nantes, France. p. 12.
- Despaigne, F., Walczak, B. and Massart, D.L. 1998. Transfer of calibration of near infrared spectra using neural networks. *Appl. Spectrosc.*, 52:732-745.
- Devaux, M.F., Robert, P., Qannary, A., Safar, M. and Vigneau, E. 1993. Canonical correlation analysis of mid and near infrared oil spectra. *Appl. Spectrosc.* 47:1024-1029.
- Dhanoa, M.S., Lister, S.J., Sanderson, R. and Barnes, R.J. 1994. The link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) transformations of NIR spectra. *J. Near Infrared Spectrosc.*, 2:43-47.
- Dhanoa, M.S., Lister, S.J. and Barnes, R.J. 1995. On the Scales Associated with Near-Infrared Reflectance Difference Spectra. *Appl. Spectrosc.*, 49:765-772.
- Díaz, I., García-Regueiro, J.A., Casillas, M. and De Pedro, E. 1996. Triglyceride composition of fresh ham fat from Iberian pigs produced with different systems of animal nutrition. *Food Chem.*, 55:383-387.
- Dobao, M.T., Rodríguez, J., Silió, L., Toro, M.A., De Pedro, E.J. and García de Siles, J.L., 1987. Crecimiento y características de canal en cerdos Ibéricos, Duroc-Jersey X Ibérico y Jiaying X Ibérico. *Inv. Agrar. Prod. Sanid. Anim.*, 2:9-23.
- Dong, J., Ma, K.M., Van de Voort, F.R. and Ismail, A.A. 1997. Stoichiometric determination of hydroperoxides in oils by fourier transform near-infrared spectroscopy. *J. AOAC Int.*, 80:345-352.
- Downey, G. 1996. Authentication of food and food ingredients by near infrared spectroscopy. *J. Near Infrared Spectrosc.*, 4:47-61.
- Downey, G. 2000. Discrimination et authentification des aliments et des ingrédients alimentaires par spectroscopie dans l'infrarouge proche et moyen. In *La spectroscopie infrarouge et ses applications analytiques*. Bertrand, D. and Dufour, E. (eds.). Editions TEC & DOC, Paris, France. p. 397-422.
- Downey, G. and Flynn, S. 2002. Discrimination between virgin olive oils from Crete, the Peleponese and other Greek islands using near infrared spectroscopy. *Near Infrared Spectroscopy: Proceedings of the 10th International Conference*. Davies, A.M.C. and Cho, R.K. NIR Publications, Chichester, UK. p. 239-241.

- El-Rafey, H.H.A., Peredi, J., Kaffka, K., Naday, B. and Balogh, A. 1988. Possibilities of application of NIR technique in the analysis of oilseeds and their derivatives. Examination of used frying oils. *Fat Sci. Tech.*, 90:175-179.
- Engelsen, S.B. 1997. Explorative spectrometric evaluations of frying oil deterioration. *J. Am. Oil Chem. Soc.*, 74:1495-1508.
- Enser, M. 1995. Meat Lipids. En *Developments in oils and fats*. Hamilton, R.J. (ed.). Blackie Academic and Professional, Chapman and Hall, London, UK. p. 1-31.
- Esbensen, K., Schönkopf, S. and Midtgaard, T. 1994. *Multivariate analysis in practice*. CAMO A/S, Trondheim, Norway.
- Fallola, A., Sabio, P., Benito, J., Ferrera, J.L. y Espárrago, C. 1989. Caracterización de la grasa externa en jamones de cerdo Ibérico. *Production porcine en Europe Méditerranéenne*. Ajaccio. Portugal.
- Fearn, F.R.B. 1982. The use of the InfrAlyzer 400L for the determination of iodine value in edible oils. *Proceedings European Launch of the Technicon InfrAlyzer 500*. Technicon Instruments Ltd.
- Fearn, T. 1986. Some statistical comments on the errors in NIR calibrations. *Analytical Proceedings*, 23:123-125.
- Fearn, T. 1996. Comparing standard deviations. *NIR News* 7, 5:5-6.
- Fenton, A.J. Jr. and Crisler, R.O. 1959. Determination of cis unsaturation in oils by near infrared spectroscopy. *J. Am. Oil Chem. Soc.*, 36:620-623.
- Fernández-Cabanás, V and Garrido-Varo, A. 1999. The use of "standard normal variate" and "detrending" as signal improvement tools for the interpretation of near infrared spectra of agro-food products. *Química Analítica*, 18:113-118.
- Firestone, D. and Horwitz, W. 1979. IUPAC Gas Chromatographic method for determination of fatty acid composition: Collaborative study. *Journal of AOAC* 62:709-721.
- Flinn, P. and Batten, G., 1995. *Leaping ahead with near infrared spectroscopy*. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia.
- Flinn, P.C. and Saunders, L.D. 1995. En *Leaping ahead with near infrared spectroscopy*. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia. p. 129.
- Flores, J. 1992. Jamones y lomos de cerdo Ibérico. Productos genuinos de la dehesa. En *Jornadas técnicas sobre obtención de productos ganaderos naturales en el ecosistema de la dehesa*. Zafra, Badajoz, España. p. 129-140.
- Flores, J., Biron, C., Izquierdo, L. and Nieto, P. 1988. Characterization of green hams from Iberian pigs by fast analysis of subcutaneous fat. *Meat Sci.*, 23:253-262.
- García-Olmo, J., Salas, J., Santaolalla, M., Peña, F. and Garrido, A. 2001b. Detección y cuantificación de adulteraciones de aceite de oliva con aceite de semillas mediante espectroscopía de infrarrojo cercano (NIRS). En *Actas del IX simposium científico-técnico Expoliva 2001*. Fundación del Olivar (ed.). Jaen, España.

- García-Regueiro, J.A. 1992. Análítica de la carne y productos del cerdo Ibérico. En El cerdo Ibérico, la naturaleza, la dehesa. Ministerio de Agricultura, Pesca y Alimentación (ed.). Badajoz, España. p. 225-234.
- Garrido A., Guerrero J.E. y Gómez Cabrera A., 1993. Application of Near Infrared Reflectance Spectroscopy (NIRS) for evaluation of animal feedstuffs. 44 th Animal Meeting of the EAAP. Aarhus. Denmark 16 -19 Agosto.
- Garrido, A., Gómez, A., Guerrero, J.E. y Fernández, V., 1996. NIRS: Una tecnología de apoyo para un servicio integral en Alimentación Animal. En Avances en Alimentación y Nutrición Animal. De Blas, C., Mateos, G.G. and Rebollar, P.G. (eds.). FEDNA, Madrid, España. p. 275-300.
- Garrido, A., Cobo, C., García-Olmo, J., Sánchez-Pineda, M.T., Alcalá, R., Horcas, J.M. and Jiménez, A. 2000. The feasibility of near infrared spectroscopy for olive oil quality control. En Near infrared spectroscopy: Proceedings of the 9th International Conference. Davies, A.M.C. and Giangiacomo, R. (eds.). NIR Publications, Chichester, UK. p. 867-871.
- Garrido, A., Pérez-Marín, M.D., Gómez, A., Guerrero, J.E., De Paz, F. and Delgado, N. 2002. NIRS analysis of molasses and fats used at the compound feedingstuffs industry. Near Infrared Spectroscopy: Proceedings of the 10th International Conference. Davies, A.M.C. and Cho, R.K. NIR Publications, Chichester, UK. p. 313-317.
- Garrido, A., García-Olmo, J. and Pérez-Marín, M.D. En prensa. Applications in the analysis of fats and oils. En NIR Spectroscopy in agriculture. Roberts, C., Workman, J. and Reeves, J. (eds.). ASA-CSSA-SSSA, Madison, WI, USA.
- Geladi P., 1996. An overview of multivariate spectral data analysis. En Near Infrared Spectroscopy: The Future Waves. Davies, A.M.C. and Williams, P.C. (eds.). NIR Publications, Chichester, UK p. 165-173.
- Geladi, P., McDougall, D. and Martens, H. 1985. Linearisation and scatter correction for near infrared reflectance spectra of meat. *Appl. Spectrosc.*, 39:491-500.
- Goddu, R.F. 1957. Determination of unsaturation by near infrared spectrophotometry. *Anal. Chem.*, 29:1790-1794.
- Goddu, R.F. 1958. Determination of phenolic hydroxyl by near infrared spectrophotometry. *Anal. Chem.*, 30:2009-2013.
- González, E. 1997. Contribución al estudio de los piensos utilizados en el cebo del cerdo Ibérico. Tesis Doctoral. Universidad de Extremadura.
- González-Martín, I., González-Pérez, C., Hernández-Méndez, J., Recio-Hernández, C. y Sabio-Rey, E. 1997. Método para la caracterización y diferenciación del cerdo Ibérico en función de la dieta. Patente española de invención N° P9502032.
- González-Martín, I., González-Pérez, C., Hernández-Méndez, J., Marqués-Macias, E. and Sanz-Poveda, F. 1999. Use of isotope analysis to characterize meat from Iberian breed swine. *Meat Sci.*, 52:437-441.
- González-Martín, I., Pérez-Pavón, J.L., González-Pérez, C., Hernández-Méndez, J. and Álvarez-García, N. 2000. Differentiation of products derived from Iberian breed swine by electronic olfactometry (electronic nose). *Anal. Chim. Acta*, 424:279-287.

- González-Martín, I., González-Pérez, C., Hernández-Méndez, J., and Sánchez-González, C. 2001. Differentiation of dietary regimene of Iberian swine by means of isotopic anlisys of carbon and sulphur in hepatic tissue. *Meat Sci.*, 58:25-30.
- Gray, J.L. and Pearson, A.M. 1987. Rancidity and warmed-oner flavors. En *Advances in meat research. Restructurated meat and poultry products*. Van Nostrand Reinhold Company. New York, NY, USA. p. 221-269.
- Ha, J., Koo, M. and Ok, H. 1998. Determination of the constituent of sesame oil by near infrared spectroscopy. *J. Near Infrared Spectrosc.*, 6:A371-A373.
- Hermida, M., Gonzalez, J.M., Sánchez, M. and Rodriguez-Otero, J.L. 2001. Moisture, solids-non-fat and fat analysis in butter by near infrared spectroscopy. *Int. Dairy Journal*, 11:93-98.
- Herschel, W. (1800). Esperiments on the refrangibility of the invisible rays of the sun. *Phil. Trans. Roy. Soc. London, Part II* 284-292.
- Hervás, C., Garrido, A., Lucena, B., Garcia, N. and De Pedro, E. 1994. Near infrared spectroscopy of Iberian pig carcasses using an artificial neural network. *J. Near Infrared Spectrosc.*, 2:177-184.
- Hildrum, K.I., Isaksson, T., Naes, T. and Tandenberg, A., 1992. Near infra-red spectroscopy. Bridging the Gap between Data Analysis and NIR Applications. Hildrum, K.I., Isaksson, T, Naes, T. and Tandberg, A. (eds.). Ellis Horwood, Chichester, UK.
- Holman, R.T. and Edmondson, P.R. 1956. Near infrared spectra of fatty acid and some related substances. *Anal. Chem.*, 28:1533-1538.
- Holman, R.T., Nickell, C. and Privett, O.S. 1958. Detection and measurement of hydroperoxides by near infrared spectrophotometry. *J. Am. Oil Chem. Soc.*, 35:422-425.
- Holman, R.T., Ener, S. and Edmondson, P.R. 1959. Detection and measurement of cis unsaturation in fatty acids. *Arch. Biochem. Biophys.*, 80:72-79.
- Hong, J.H., Yamaoka-Koseki, S. and Yasumoto, K. 1994. Analysis of peroxide values in edible oils by near infrared spectroscopy. *Nippon Shokuhin Kogyo Gakkaishi* 41:277-280.
- Hourant, P., Baeten, V., Morales, M.T., Meurens, M. and Aparicio, R. 2000. Oil and fat classification by selected bands of Near-Infared Spectroscopy. *Appl. Spectrosc.* 54:1168-1174.
- Hruschka, W.R. 1987. Data Analysis: Wavelength selection methods. En *Near Infrared Technology in the Agricultural and Food Industries*. William, P. and Norris, K. (eds.). American Association of Cereal Chemist Inc., St. Paul, Minnesota, USA. p. 35-55.
- Indahl, U.G., Sahni, N.S., Kirkhus, B. and Naes, T. 1999. Multivariate strategies for classification based on NIR spectra with application to mayonnaise. *Chem. Intell. Lab. Syst.*, 49:19-31.
- Isaksson, T. and Naes, T. 1990. Selection of samples for calibration in near infrared spectroscopy. Part II: Selection based on spectral measurement. *Appl. Spectrosc.*, 44,1152-1158.

- ISO 5725-2. 1994. Accuracy (trueness and precision) of measurement methods and results. Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. International Organization for Standardization (ISO), Geneva, Switzerland.
- IUPAC n° 2301. 1987. Preparation of the fatty acid methyl esters. Standard Methods for the analysis of oils, fats and derivatives. International Union of Pure and Applied Chemistry, Blackwell, Oxford, UK.
- Iwamoto, M., 1991. Application of Near Infrared Spectroscopy for quality control in food and feed industries. En Proceedings of the 3rd International Conference on Near Infrared Spectroscopy. Biston, R. and Bartiaux-Thill, N. (eds.). Agricultural Research Centre, Gembloux, Belgium. p. 175-186.
- Iwamoto, M. and Kawano, S. 1990. Proceedings of the 2nd International Conference. Iwamoto, M. and Kawano S. (eds.). Koring Publishing Ltd, Japan.
- Izquierdo, L. y Nieto, P. 1989. Caracterización de grasas de cerdo Ibérico con distintos tipos de alimentación. Avances en la tecnología del jamón curado. En II Jornadas Técnicas sobre el jamón curado. Valencia, España. p. 65-85.
- Kaffka, K. and Giarmati, L.S. 1995. Quality determination in the food industry. En Leaping ahead with near infrared spectroscopy. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group, Royal Australian Chemistry Institute, Melbourne, Australia. p. 261-269.
- Kawano, S. and Iwamoto, M. 1995. Overview of novel applications of near infrared spectroscopy for the food industry in Japan. En Leaping ahead with near infrared spectroscopy. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group, Royal Australian Chemistry Institute, Melbourne, Australia. p. 270-276.
- Kramer, R. 1998. Chemometric techniques for quantitative analysis. Marcel Dekker Inc., New York, NY, USA.
- León, F. 1992. Optimización de los parámetros de aceptabilidad del cerdo Ibérico. En: El cerdo Ibérico, la naturaleza, la dehesa. Ministerio de Agricultura, Pesca y Alimentación (ed.). Badajoz, España. p. 245-255.
- Li, H., Van de Voort, F.R., Sedman, J. and Ismail, A.A. 1999. Rapid determination of cis and trans content, iodine value, and saponification number of edible oils by fourier transform near-infrared spectroscopy. J. Am. Oil Chem. Soc., 76:491-497.
- López-Bote, C.J. 1998. Sustained utilization of the Iberian pig breed. Meat Sci., 49:s17-s27.
- Lu, J.; McClure, W.F., Barton, F.E. and Himmelsbach, D.S. 1998. Effect of random noise on the performance of NIR calibrations. J. Near Infrared Spectrosc. 6:77-87.
- Mahalanobis, P.C. 1936. On the generalised distance in statistics. Proceedings of the National Institute of Science of India, 12:49-55.
- Man, Y.B.C. and Moh, M.H. 1998. Determination of free fatty acids in palm oil by near-infrared reflectance spectroscopy. J. Amer. Oil Chem. Soc., 75:557-562.

- MAPA. 1984. Una imagen de calidad. Los productos del cerdo Ibérico. Ministerio de Agricultura, Pesca y Alimentación (ed.). MAPA, Madrid, España.
- Mark, H. 1992. Qualitative discriminant analysis. En Handbook of near-infrared analysis. Burns, D.A. and Ciurczak, E.W. (eds.). Marcel Dekker, New York, NY, USA. p. 329-363.
- Mark, H. and Workman, Y. 1991. Statistics in Spectroscopy. Academic Press Inc., New York, NY, USA.
- Martens, H. and Naes, T. 1989. Multivariate calibration. John Wiley and Sons Inc., Chichester, UK.
- Martens, H. and Martens, M. 2000. Multivariate analysis of quality: an introduction. John Wiley and Sons Inc., Chichester, UK.
- Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. and Kaufman, L. 1988. Chemometrics: a textbook. Elsevier Science Pub., Amsterdam, The Netherlands.
- Massart, D.L., Vandeginste, B.G.M., Buydens, L.M.C., De Jong, S., Lewi, P.J. and Smeyers-Verdeke, J. 1997. Handbook of Chemometrics and Qualimetrics (2 vol.). Elsevier Science Pub., Amsterdam, The Netherlands.
- McClure, W.F. 1987. Near Infrared instrumentation. En Near-infrared technology in the agricultural and food industries. Williams, P.C. and Norris, K. (eds.). American Association of Cereal Chemists Inc., St. Paul, MN, USA. p. 89-105.
- McClure, W.F. 1992b. Making light work: lighting new frontiers. En Making light work: advances in near infrared spectroscopy. Murray, I. and Cowe, I.A. (eds.). VCH. London, UK. p. 1-13.
- McElhinney, J., Downey, G. and Fearn, T. 1999. Chemometric processing of visible and near infrared reflectance spectra for species identification in selected raw homogenised meats. *J. Near Infrared Spectrosc.* 7:145-154.
- Mitsumoto, M., Maeda, S., Mitsuhashi, T. and Ozawa S. 1991. Near Infrared Spectroscopy Determination of physical and chemical characteristics in beef cuts. *J. Food Sci.*, 56:1493-1496.
- Morrison, D.F. 1976. Multivariate statistical methods (2nd ed.). Mac Graw Hill Inc., NY, USA.
- Moya, L. 1993. Aplicación de la tecnología NIRS al control de calidad en la industria de piensos compuestos. Trabajo Profesional Fin de Carrera. ETSIAM. Universidad de Córdoba.
- Moya, L., Garrido, A., Guerrero, J.E., Lizaso, J. and Gómez, A., 1995. Quality control of raw materials in the feed compounds industry. En Leaping ahead with near infrared spectroscopy. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group, Royal Australian Chemistry Institute, Melbourne, Australia. p. 111-117.
- Murray, I. 1987. The NIR spectra of homologous series of organic compounds. En NIR/NIT Spectroscopy. Hollo, J., Kaffka, K.J. and Gonczy, J.L. (eds.). Akademiai Kiado, Budapest, Hungary. p. 13-28.

- Murray, I. 1990. Applications of NIRS in Agriculture. En Proceedings of the 2nd International near infrared spectroscopy conference. Iwamoto, M. and Kawano, S. (eds.). Korian Publishing Co, Japan. p. 11-20.
- Murray, I. and Williams, P.C. 1987. Chemical principles of near infrared technology. En Near infrared technology in the agricultural and food industries. Williams, P.C. and Norris, K. (eds.). American Association of Cereal Chemists, St. Paul, Minnesota, USA. p. 17-34.
- Murray, I. and Cowe, I. 1992. Making Light Work: Advances in Near Infrared Spectroscopy. Murray, I. and Cowe, I.A. (eds.). VCH. London, UK.
- Naes, T. and Isaksson, T. 1989. Selection of samples for near infrared spectroscopy. Part I: General principles illustrated by examples. *Appl. Spectrosc.*, 43:328-335.
- Naes, T., Isaksson, T., Fearn, T. and Davies, T. 2002. A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK.
- Ordoñez, J.A. y De la Hoz, L. 1992. Alimentación y calidades de carnes del cerdo Ibérico. En: El cerdo Ibérico, la naturaleza, la dehesa. Ministerio de Agricultura, Pesca y Alimentación. (ed.). Badajoz, España. p. 209-221.
- Osborne, B.G. and Fearn, T. 1986. Near Infrared Spectroscopy in Food Analysis. Longman Scientific and Technical, Halow, Essex, UK.
- Osborne, B.G., Fearn, T. and Hindle, P. 1993. Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Longman Scientific and Technical. London, UK.
- Osborne, B.G., Kotwal, Z., Wesley, I.J., Saunders, L., Dardenne, P. and Shenk, J. 1999. Optical matching of near infrared reflectance monochromator instruments for the analysis of ground and whole wheat. *J. Near Infrared Spectrosc.*, 7:167-178.
- Osorio, E., Montero, V., Fallola, A., Benito, J., Ferrera, J.L. y Sabio, E. 1992. Composición en ácidos grasos de la grasa subcutánea de cobertura en jamones frescos de cerdos de tronco Ibérico con distintos tipos de alimentación. En: II Coloquios sobre el Cerdo Mediterraneo. Badajoz, España. p. 83-84.
- Otto, M. 1998. Chemometrics: statistics and computer application in analytical chemistry. Wiley-VCH, Weinheim, Germany.
- Panford, J.A. and De Man, J.M. 1990. Determination of oil content of seeds by NIR. Influence of fatty acid composition on wavelength. *J. Am. Oil Chem. Soc.*, 67:473-482.
- Park, R.S., Agnew, R.E. and Barnes, R.J. 1999. The development of near infrared reflectance spectroscopy calibrations for undried grass silage and their transfer to another instrument using multiple and single sample standardisation. *J. Near Infrared Spectrosc.*, 7:117-131.
- Peralta, M.A. 1996. Caracterización del jamón: influencia de la raza y alimentación en los ácidos grasos y compuestos volátiles de la grasa. Tesis Doctoral. Universidad de Córdoba.
- Pérez-Marín, M.D., De Pedro, E., García-Olmo, J. and Garrido, A. 2002. NIRS analysis of Iberian pig fat: repeatability file effect. Near Infrared Spectroscopy: Proceedings of the 10th International Conference. Davies, A.M.C. and Cho, R.K. NIR Publications, Chichester, UK. p. 473-475.

- Pérez-Vich, B., Velasco, L. and Fernández-Martínez, J.M. 1998. Determination of seed oil content and fatty acid composition in sunflower through the analysis of intact seeds, husked seeds, meal and oil by near infrared reflectance spectroscopy. *J. Am. Oil Chem. Soc.*, 75:547-555.
- Puigdoménech, A. 1998. Optimització del control de qualitat en productes agrícoles mitjançant espectrometria d'infraroig proper (NIR): aplicació als farratges de Catalunya. Transferència d'equacions de calibratge multivariant en xarxes instrumentals. Tesis Doctoral. Universitat de Barcelona.
- Puigdoménech, A., Garrido, A., De la Roza, B., Aragay, M., Lucena, B. and Martínez, A. 1995. Harmonization of animal feeds evaluation using NIRS. En *Leaping ahead with near infrared spectroscopy*. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia. p. 107-110.
- Rey, A.I., López-Bote, C.J. and Arias R.S. 1997. Effect of extensive feeding on alpha-tocopherol concentration and oxidative stability of muscle microsomes from Iberian pigs. *Animal Sci.*, 65:515-520.
- Rey, A.I., Isabel, B., Cava, R. and López-Bote, C.J. 1998. Dietary acorns provide a source of gamma-tocopherol to pigs raised extensively. *Can. J. of Anim. Sci.*, 78:441-443.
- Riaublanc, A., Bertrand, D. and Dufour, E. 2000. Lipides. En *La spectroscopie infrarouge et ses applications analytiques*. Bertrand, D. and Dufour, E. (eds.). Editions TEC & DOC, Paris, France. p. 139-166.
- Ripoche, A. and Guillard, A.S. 2001. Determination of fatty acid composition of pork fat by fourier transform infrared spectroscopy. *Meat Sci.*, 58:299-304.
- Robert, P., Devaux, M.F. and Safar, M. 1995. Canonical correlation analysis: application to near and mid infrared spectroscopy. En *Leaping ahead with near infrared spectroscopy*. Batten, G.D., Flinn, P.C., Welsh, L.A. and Blakeney, A.B. (eds.). NIR Spectroscopy Group. Royal Australian Chemistry Institute, Melbourne, Australia. p. 62-65.
- Roggo, Y., Duponchel, L., Noe, B. and Huvenne, J.P. 2002. Sucrose content determination of sugar beets by near infrared reflectance spectroscopy. Comparison of calibration methods and calibration transfer. *J. Near Infrared Spectrosc.*, 10:137-150.
- Rosales M. 1993. Uso de la Información Espectroscópica NIRS "per se" en la Evaluación de la Calidad de Productos Agrícolas. Trabajo Profesional Fin de Carrera. ETSIAM. Universidad de Córdoba.
- Ruiz, J., Cava, R., Antequera, T., Martín, L., Ventanas, J. and López-Bote, C.J., 1998. Prediction of the feeding background of Iberian pigs using the fatty acid profile of subcutaneous, muscle and hepatic fat. *Meat Sci.*, 49:155-163.
- Ruiz, J. and Petró, M.J. 2000. Métodos para la clasificación de la materia prima. En *Tecnología del jamón Ibérico. De los sistemas tradicionales a la explotación racional del sabor y el aroma*. Ventanas, J. (coord.). Ediciones Mundi-Prensa, Madrid, España. p.131-160.
- Salas, J. Comunicación personal. Laboratorio Agroalimentario Córdoba. Consejería de Agricultura y Pesca. Junta de Andalucía.

- SAS. 1990. SAS/STAT User's Guide Version 6. SAS Institute Inc. (ed.). Cary, NC, USA.
- Sasaki, M. and Ozawa, S. 1990. Determination of the ingredients of meat and meat products by Near infrared spectroscopy. En Proceedings of the 2nd International near infrared spectroscopy conference. Iwamoto, M. and Kawano, S. (eds.). Korian Publishing Co, Japan. p. 385- 392.
- Sato, T. 1994. Application of principal component analysis on near infrared spectroscopic data of vegetable oils for their classification. J. Am. Oil Chem. Soc., 71:293-298.
- Sato, T., Yoshino, M. and Cho, R.K. 1988. Near infrared reflectance spectroscopic analysis for butter constituents. Jpn. J. Zootech. Sci., 59:806-813.
- Sato, T., Yoshino, M., Suzuki, I., Kawano, S. and Iwamoto, M. 1989. Use of near infrared spectroscopy to detect foreign fat adulteration of dairy products. En Proceedings of the 2nd International near infrared spectroscopy conference. Iwamoto, M. and Kawano, S. (eds.). Korian Publishing Co, Japan. p. 148-156.
- Sato, T., Kawano, S. and Iwamoto, M. 1990. Detection of foreign fat adulteration of milk fat by near infrared spectroscopy method. J. Dairy Res., 73:3408-3413.
- Sato, T., Kawano, S. and Iwamoto, M.. 1991. Near infrared spectral patterns of fatty acid analysis from fats and oils. J. Am. Oil Chem. Soc., 68:827-833.
- Sato, T., Takahata, Y., Noda, T., Yanagisawa, T., Morishita, T. and Sakai, S. 1995. Nondestructive determination of fatty acid composition of husked sunflower (*Helianthus annua* L.) seeds by near infrared spectroscopy. J. Am. Oil Chem. Soc., 72:1177-1183.
- Shenk, J.S. 1985. En Near Infrared Reflectance Spectroscopy (NIRS): Analysis of forage quality. Marten, G.C., Shenk, J. and Barton, F.E. (eds.). USDA. ARS. Agriculture Handbook N° 643, Washington, USA.
- Shenk, J.S. 1990. Standardizing NIRS instruments. En Proceedings of the 3rd International Conference on Near Infrared Spectroscopy. Biston, R. and Bartiaux-Thill, N. (eds.). Agricultural Research Centre, Gembloux, Belgium. p. 649-654.
- Shenk, J.S. 1992. Networking and calibration transfer. En Making light work: advances in near infrared spectroscopy. Murray, I. and Cowe, I.A. (eds.). VCH. London, UK. p. 223-228.
- Shenk, J.S. and Westerhaus, M.O. 1989. U.S. Pat., 4866644, Sept 12 1989.
- Shenk, J.S., Westerhaus, M.O. and Abrams, S.M. 1989. Protocols for NIRS calibration: monitoring analysis results and recalibration: En Near Infrared Reflectance Spectroscopy (NIRS): Analysis of forage quality. Marten, G.C., Shenk, J. and Barton, F.E. (eds.). USDA. ARS. Agriculture Handbook N° 643, Washington, USA. p. 104-110.
- Shenk, J.S and Westerhaus, M.O. 1991a. Population definition sample selection and calibration procedures for near infrared spectra and modified partial least squares regression. Crops Sci., 31:469-474.
- Shenk, J.S. and Westerhaus, M.O. 1991b. New standardisation and calibration procedure for NIRS Analytical Systems. Crop Sci., 31:1694-1696.

- Shenk, J.S., Workman, J.J. and Westerhaus, M.O. 1992. Application of NIR Spectroscopy to Agricultural Products. En Handbook of near infrared analysis. Burns, P.A. and Ciurczak, E.W. (eds.). Marcel Dekker, New York, NY, USA. p. 383-427.
- Shenk, J.S. and Westerhaus, M.O. 1995a. Analysis of Agricultural and Food Products by Near Infrared Reflectance Spectroscopy. Monograph. NIRSystems Inc., Silver Spring, MD, USA.
- Shenk, J.S. and Westerhaus, M.O. 1995b. Routine operation, calibration, development and network system management manual. NIRSystems Inc., Silver Spring, MD, USA.
- Shenk, J.S. and Westerhaus, M.O. 1996. Calibration the ISI way. En Near Infrared Spectroscopy: The Future Waves. Davies, A.M.C. and Williams, P.C. (eds.). NIR Publications, Chichester, UK. p. 198-202.
- Sinnaeve, G., Dardenne, P. and Biston, R.. 1990. Predicting of fat, moisture and solids-not-fat content of butter by near infrared reflectance spectroscopy. Proceedings of the 3rd International Conference on Near Infrared Spectroscopy. Biston, R. and Bartiaux-Thill, N. (eds.). Agricultural Research Centre, Gembloux, Belgium. p. 249-257.
- Stark, E. 1996. Near Infrared Spectroscopy past and future. En Near Infrared Spectroscopy: The Future Waves. Davies, A.M.C. and Williams, P.C. (eds.). NIR Publications, Chichester, UK. p. 700-713.
- Stone, M. 1974. Cross-validation choice and assessment of statistical prediction. J. R. Statist. Soc. B, 39:111-133.
- Tejeda, J.F. 1999. Estudio de la influencia de la raza y la alimentación sobre la fracción lipídica intramuscular del cerdo Ibérico. Tesis Doctoral. Universidad de Extremadura.
- Tejeda, J.F., García, C, Petró, M.J., Andrés, A.I. and Antequera, T. 2001a. n-Alkane content of intramuscular lipids of Iberian fresh ham from different feeding systems and crossbreeding. *Meat Sci.*, 57:371-377.
- Tejeda, J.F., Antequera, T., Martín, L., Ventanas, J. and García, C. 2001b. Study of the branched hydrocarbon fraction of intramuscular lipids from Iberian fresh ham. *Meat Sci.*, 58:175-179.
- Tejeda, J.F., García, C, Antequera, T. y Ventanas, J. 2001c. Método de control analítico de la carne de cerdo. Solicitud de Patente española N° P9901519.
- Tejeda, J.F., Gandemer, G., Antequera, T., Viau, M. and García, C. 2002. Lipid traits of muscles as related to genotype and fattening diet in Iberian pigs: total intramuscular lipids and triacylglycerols. *Meat Sci.*, 60:357-363.
- UNE 5508. 1996. Aceites y grasas de origen animal y vegetal. Análisis por cromatografía en fase gaseosa de los ésteres metílicos de ácidos grasos. Asociación española de Normalización y Certificación (AENOR), Madrid, España.
- UNE 82009-2. 1999. Exactitud (veracidad y precisión) de resultados y métodos de medición. Parte 2: Método básico para la determinación de la repetibilidad y la reproducibilidad de un método de medición normalizado. Asociación española de Normalización y Certificación (AENOR), Madrid, España.

- Valdés, E.V. and Leeson, S. 1994. Measurement of ME, GE and moisture in feed grade fats. *Poultry Sci.*, 71:1559-1563.
- Ventanas, J. 2000. Tecnología del jamón Ibérico. De los sistemas tradicionales a la explotación racional del sabor y el aroma. Ediciones Mundi-Prensa, Madrid, España.
- Vidal-Aragón, M.C. 1996. Influencia de la materia prima en la calidad y en el aroma del jamón curado de cerdo Ibérico. Tesis Doctoral Universidad de Extremadura.
- Wang, Y., Velkamp, D. and Kowalski, B.R. 1991. Multivariate instrument standardization. *Anal. Chem.*, 63:2750-2756.
- Weaver, R.W.V. 1984. En Challenge to contemporary dairy analytical techniques. The Royal Society of Chemistry, Burlington House, London, UK. p. 91-102. (referenced by Sato *et al.*, 1988).
- Wesley, I.J., Barnes, P.J. and McGuill, A.E.J. 1995. Measurements of adulteration of olive oils by near infrared spectroscopy. *J. Am. Oil. Chem. Soc.*, 72:289-292.
- Wesley, I.J., Pacheco, F. and McGill, A.E.J. 1996. Identification of adulterants on olive oils. *J. Am. Oil Chem. Soc.*, 73:515-518
- Wetzel, D.L. 1983. Lipid structures and NIRA. En Proceedings. Int. 4th Symp. Near Infrared Reflectance Analysis. Technicon Industrial Systems. Tarrytown, NY, USA.
- Williams, P.C. 1987. Variables affecting near-infrared reflectance spectroscopy analysis. En Near-infrared technology in the agricultural and food industries. Williams, P.C. and Norris, K. (eds.). American Association of Cereal Chemists Inc., St. Paul, MN, USA. p. 143-148.
- Williams, P.C. and Norris, K. 1987a. Near-infrared technology in the agricultural and food industries. Williams, P.C. and Norris, K. (eds.). American Association of Cereal Chemists Inc., St. Paul, MN, USA.
- Williams, P.C. and Norris, K. 1987b. Qualitative applications of Near-Infrared Reflectance Spectroscopy. In P.C. Williams and K. Norris (ed.) En Near-infrared technology in the agricultural and food industries. Williams, P.C. and Norris, K. (eds.). American Association of Cereal Chemists Inc., St. Paul, MN, USA. p. 241-246.
- Williams, P.C. and Sobering, D. 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.*, 1:25-32.
- Williams, P.C. and Sobering, D., 1996. How do we do it: a brief summary of the methods we use in developing near infrared calibrations En Near Infrared Spectroscopy: The Future Waves. Davies, A.M.C. and Williams, P.C. (eds.). NIR Publications, Chichester, UK. p. 185-188.
- Windham, W.R., Martens, D.R. and Barton, F.E. 1989. Protocols for NIRS calibration: sample selection and equation development and validation. En Near Infrared Reflectance Spectroscopy (NIRS): Analysis of forage quality. Marten, G.C., Shenk, J. and Barton, F.E. (eds.). USDA. ARS. Agriculture Handbook N° 643. Washington, USA. p. 96-103.

- Windham, W.R., and Flinn, P.C. 1992. Comparison of MLR and PLS regression in NIR analysis of quality components in diverse feedstuff populations. En *Near infrared spectroscopy. Bridging the Gap between Data Analysis and NIR Applications*. Hildrum, K.I., Isaksson, T, Naes, T. and Tandberg, A. (eds.). Ellis Horwood, Chichester, UK. p. 459-464.
- Wold, S. 1972. Spline-funktioner-et nytt verktyg i data-analysen. *Kemisk Tidskr.*, 3:34-37.
- Wold, S. 1976. Pattern recognition by means of disjoint principal components models. *Pattern Recogn.*, 8:127-139.
- Wold, S. and Sjöström, M. 1998. Chemometrics, present and future success. *Chem. Intell. Lab. Syst.*, 44:3-14.
- Wold, S., Antii, S.H., Lindgren, F. and Öhman, J. 1998. Ortogonal signal correction of near infrared spectra. *Chem Intell. Lab. Syst.* 44:175-185.
- Yildiz, G., Wehling, R. and Cuppet, S. 2001. Method for determining oxidation of vegetable oils by Near-Infrared Spectroscopy. *J. Am. Oil. Chem. Soc.*, 78:495-502.
- Yoshida, H. 1993. Influence of fatty acids of different unsaturation in the oxidation of purified vegetable oils during microwave irradiation. *J. Sci. Food Agric.*, 64:41-47.
- Zhang, H.Z. and Lee, T.C. 1997. Rapid near infrared spectroscopic method for the determination of free fatty acid in fish and its application in fish quality assessment. *J. Agric. Food Chem.* 45:3515-3521.