

UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE

ESCUELA POLITÉCNICA SUPERIOR DE ORIHUELA

GRADO EN INGENIERÍA AGROALIMENTARIA Y AGROAMBIENTAL



**UTILIZACIÓN DE LA TÉCNICA NIRS PARA EVALUAR
LA COMPOSICIÓN QUÍMICA DE SUBPRODUCTOS
VEGETALES DESTINADOS EN ALIMENTACIÓN ANIMAL**

TRABAJO FIN DE GRADO

Julio 2016

Autor: Francisco Javier Díaz Campos

Tutor/es: José Ramón Díaz Sánchez
Raquel Muelas Domingo



UNIVERSITAS

Miguel

Hernández

RESUMEN

Para analizar la composición nutricional de muestras de subproductos vegetales destinados para la alimentación animal se han desarrollado modelos de estimación para la grasa bruta, proteína bruta, fibra bruta, materia seca, materia orgánica y cenizas usando un equipo de espectroscopía en el infrarrojo cercano (NIRS). Para ello se utilizaron 100 muestras de diversos subproductos vegetales y se aplicó el algoritmo PLS (mínimos cuadrados) con distintos pretratamientos espectrales para el desarrollo de los modelos de estimación. El modelo óptimo de estimación para la grasa bruta, proteína bruta, fibra bruta, materia seca, materia orgánica y cenizas presentó coeficientes de determinación (R^2_{cv}) de 0,749; 0,939; 0,794; 0,931; 0,930 y 0,902 respectivamente.

Palabras clave: NIRS, subproductos vegetales, alimentación animal, composición química, medio ambiente.

ABSTRACT

To analyze the nutritional composition of vegetable by-products destined for animal feed, estimation models were developed for crude fat, crude protein, crude fibre, dry matter, organic matter and ash using near infrared spectroscopy (NIRS). One hundred samples of different vegetable by-products were used. Partial least squares (PLS) regression was applied with different pretreatment of spectra for model development. The optimal estimation for crude fat, crude protein, crude fibre, dry matter, organic matter and ash had coefficients of correlation (R^2_{cv}) de 0,749; 0,939; 0,794; 0,931; 0,930 and 0,902 respectively.

Keywords: NIRS, vegetable by-products, animal feed, chemical composition,

AGRADECIMIENTOS

Quisiera expresar mi gratitud a mis tutores, José Ramón Díaz por ofrecerme la oportunidad trabajar con él, en este proyecto, de gran futuro, quiero destacar muy especialmente a Raquel Muelas que ha sido la artífice de que yo pudiera realizar este trabajo y llegar a buen término ya que estuvo en cada momento, y en cada duda, sin ella no hubiera sido posible.



INDICE

1. INTRODUCCIÓN	7
1.1. IMPORTANCIA EN LA PRODUCCIÓN DE SUBPRODUCTOS.....	7
1.2 UTILIZACIÓN DE SUBPRODUCTOS EN LA ALIMENTACIÓN ANIMAL.	9
1.2.1 SUBPRODUCTOS UTILIZADOS COMO TAL	9
1.2.2 SUBPRODUCTOS UTILIZADOS COMO ENSILADO.	10
1.2.3 EL PROCESO DE ENSILAJE.	10
1.2.4 CALIDAD DE LOS ENSILADOS.	12
1.3 APLICACIÓN DE LA TECNOLOGÍA NIRS EN EL ANALISIS DE SUBPRODUCTOS Y ENSILADOS.	16
1.3.1. ESPECTROCOPIA NIR.	17
1.3.2. INTRODUCCIÓN A LA ESPECTROCOPIA NIR.....	17
1.3.4 BASES TEORICAS DE LA ABSORCIÓN NIR.	20
1.3.5 ESPECTRO ELECTROMAGNETICO.....	22
1.3.6 QUIMIOMETRÍA Y SU APLICACIÓN EN LA TECNOLOGÍA NIRS.....	23
1.3.7 ETAPAS DEL DESARROLLO DE MODELOS DE CALIBRACIÓN MULTIVARIABLES.	26
1.3.8 OBTENCIÓN DE LOS ESPECTROS DE ABSORCIÓN.	26
1.3.9 SELECCIÓN DEL CONJUNTO DE MUESTRAS.	26
1.3.9.1 MÉTODO DE REFERENCIA.....	26
1.3.9.2 PRETRATAMIENTO ESPECTRAL.	27
1.3.9 DESARROLLO DEL MODELO DE CALIBRACIÓN.....	30
1.3.10 EVALUACIÓN DEL MODELO DE CALIBRACIÓN.....	33
1.3.11 PARÁMETROS PARA EVALUAR EL MODELO DE CALIBRACIÓN.	33
2. OBJETIVOS	37
3. MATERIALES Y METODOS.	38
3.1. MUESTRAS Y ANÁLISIS NIRS.	38
3.2. INSTRUMENTO Y ANÁLISIS NIRS.....	38
3.2. ANÁLISIS QUÍMICOS DE REFERENCIA.	40
3.2.1. DETERMINACIÓN DE MATERIA SECA.	40
3.3.2. DETERMINACIÓN DE CENIZAS.	41
3.3.4. DETERMINACIÓN DE GRASA BRUTA.	43
3.3.5. DETERMINACIÓN DE PROTEINA BRUTA.	44
3.3.6. DETERMINACIÓN DE FIBRA BRUTA.	45

3.3.7. DESARROLLO DE LOS MODELOS DE CALIBRACIÓN Y EVALUACIÓN DE LAS ECUACIONES.....	45
4. RESULTADOS Y DISCUSIÓN	48
4.1. CARACTERISTICAS DE LAS MUESTRAS DE SUBPRODUCTOS VEGETALES UTILIZADOS PARA LA CALIBRACIÓN	48
4.2. DESARROLLO Y EVALUACIÓN DE LAS ECUACIONES DE CALIBRACIÓN SIN PRETRATAMIENTO ESPECTRAL.....	51
4.3. OPTIMIZACIÓN DE LAS ECUACIONES DE CALIBRACIÓN DESARROLLADAS Y EL ESTUDIO DE SU CAPACIDAD PREDICTIVA Y DE ESTIMACIÓN.....	55
5. CONCLUSIONES.....	66
6. BIBLIOGRAFIA.....	68



1. INTRODUCCIÓN

1.1. IMPORTANCIA EN LA PRODUCCIÓN DE SUBPRODUCTOS.

El sector agroalimentario en España genera gran cantidad de subproductos y residuos que carecen de valor económico, y que contribuyen de forma notable al deterioro ambiental, su eliminación es un costoso proceso que repercute en el incremento del precio final de los productos y esto a su vez supone un detrimento de la competitividad del sector. Los recursos utilizados en el sector de alimentación son de origen agrícola, y debido a su naturaleza biológica, prácticamente cada parte de una cosecha agrícola tiene una aplicación útil, y un potencial inmenso para utilizar estos recursos de una manera muy eficiente. En España los productos agrícolas significan actualmente más del 50% de la producción final agraria.

Según el último anuario de estadística publicado por el Ministerio de Agricultura, Alimentación y Medio Ambiente, correspondiente al año 2014 la superficie total cultivada en España supera en algo más de 45 millones de hectáreas si se desglosa por ocupación se tendría por un lado la superficie de uso principal de pastos (pastizales, prados, y eriales) que ocupa 9.6 millones de ha, por otro lado la superficie forestal arbolada, arbustiva y de matorral ocupa 19.4 millones de ha, y por último las tierras de cultivo con 17.132.780 de ha, donde su producción cuyo destino es la industria alimentaria y con predominio de los cultivos herbáceos, (cereales, leguminosas, hortícolas) con una extensión de alrededor de 9 millones de hectáreas, seguida de los cultivos leñosos cercana a las 5 millones de hectáreas (**Tabla 1**). En los cultivos hortícolas destacan entre otros, sandía, melón, tomate, calabacín, pepino coliflor, brócoli, lechuga, y se presentan tanto al aire libre como en invernadero. Por otro lado los cultivos leñosos adquieren importancia los cítricos, el olivo y la vid, de estos se obtienen fundamentalmente restos de pulpa, ramón, hojarasca, y otros restos de cosecha. Tanto en el sector agrícola primario con prácticas agrícolas y labores culturales realizadas como pueden ser la poda, destallado, deshojado, aclareo de flores y frutos, y en el sector

agrícola secundario o de transformación de producto agrícola en producto alimentario en su proceso propio de transformación industrial, a través de la separación y eliminación de partes de los frutos así como de los residuos resultantes de dicho proceso y que pueden generar entre un 15 y un 65% de subproducto dependiendo de la materia prima.

Tabla 1. Distribución general del suelo por usos y aprovechamientos en tierras de cultivo en España

Aprovechamientos	Año 2013 Miles (ha)
Cultivos herbáceos	9.105.059
Barbechos y otras tierras no ocupadas	3.076.081
Cultivos leñosos	4.951.640
TOTAL TIERRAS DE CULTIVO	17.132.780

Fuente: Anuario de Estadística. Magrama, 2014

A continuación se detallan los conceptos que se indican en las observaciones metodológicas realizadas para esta estadística:

Tierras de cultivo: Incluye tierras ocupadas por cultivos herbáceos, barbechos y otras tierras no ocupadas y tierras ocupadas por cultivos leñosos.

Tierras ocupadas por cultivos herbáceos: Engloban las tierras bajo cultivos temporales (considerando sólo una vez las que obtienen dos cosechas), las praderas temporales para siega o pastoreo, y las tierras dedicadas a huertas con inclusión de cultivos de invernadero.

Barbechos y otras tierras no ocupadas: Se incluyen en este grupo todas las tierras de cultivo en descanso o no ocupadas durante el año, independientemente del motivo, aunque hayan sido aprovechadas como pasto para el ganado.

Tierras ocupadas por cultivos leñosos: Se refieren a las tierras con cultivos que ocupan el terreno durante largos periodos y no necesitan

ser replantados después de cada cosecha. Incluye tierras ocupadas por árboles frutales, árboles de fruto seco, olivos, vides, etc., pero excluye la tierra dedicada a árboles para la producción de leña o madera.

Dada la importante producción de la industria agroalimentaria, la sostenibilidad medioambiental en la cadena alimentaria requiere de la implicación de todos los componentes del ciclo de vida, y cada integrante de la cadena tiene un papel crucial a desempeñar, de modo tanto individual como colectivo. El objetivo declarado de la industria agroalimentaria como productora de subproductos es utilizar el 100% de los recursos agrícolas reduciendo residuos al mínimo posible, cumpliendo así la normativa vigente, y transformando estos en una amplia gama de productos, incluyendo coproductos y subproductos, que pueden ser destinados a la alimentación animal.

1.2 UTILIZACIÓN DE SUBPRODUCTOS EN LA ALIMENTACIÓN ANIMAL.

1.2.1 SUBPRODUCTOS UTILIZADOS COMO TAL

El uso de los subproductos agrícolas como fuente de alimento para el ganado ha sido puesto en práctica a lo largo de la historia. Entre otros subproductos destacan restos de cosechas, podas como ramón de olivo, sarmiento de vid, el tomate, pimiento, pulpa de cítricos, fresas, bagazo y levadura de cerveza, etc., estos son abundantes a lo largo de la geografía española de ahí su importancia en su explotación para la alimentación animal. La posibilidad de emplear una amplia gama de productos de gran calidad de modo continuado durante todo el año y de fácil disponibilidad y a un bajo coste, hace que repercuta de forma notable en la rentabilidad de las explotaciones ganaderas, en la actualidad, además, del interés por su aprovechamiento no es sólo de carácter económico, sino también ecológico, los subproductos del sector alimentario mantienen la transición hacia una economía ecológica de ir reduciendo los efectos contaminantes, y de poner en valor, el uso de estos productos en la alimentación animal y que está ampliamente

regulado en la legislación de la UE, incluido en el Reglamento 178/2002 sobre legislación alimentaria.

1.2.2 SUBPRODUCTOS UTILIZADOS COMO ENSILADO.

Actualmente la industria agroalimentaria productora de subproductos para el ganado sólo ha incorporado una parte de la variedad de subproductos generados debido a su elevado contenido en agua.

En este sentido, **el ensilado** es un método de conservación común para el forraje que con ciertas modificaciones tecnológicas, también puede ser utilizado con subproductos agroalimentarios de elevado contenido en agua. El objetivo es mantener el valor nutritivo original, con un mínimo de pérdidas en materia seca.

1.2.3 EL PROCESO DE ENSILAJE.

El ensilaje se logra por medio de una fermentación láctica espontánea en condiciones anaerobias. Las bacterias epifíticas de ácido láctico

(BAC) fermentan los carbohidratos hidrosolubles (CHS) del forraje produciendo ácido láctico y en menor cantidad, ácido acético. Al generarse estos ácidos el pH del material ensilado baja a un nivel que inhibe la presencia de microorganismos que inducen la putrefacción. El proceso del ensilaje se puede dividir en cuatro etapas:

Fase 1 - Fase Aeróbica.

Esta fase dura pocas horas. El oxígeno atmosférico presente en la masa vegetal disminuye rápidamente debido a la respiración de los microorganismos aerobios y aerobios facultativos como las levaduras y enterobacterias. Además, hay actividad de varias enzimas vegetales, como las proteasas y las carbohidrasas, siempre que el pH se mantenga en el rango normal para el jugo del forraje fresco (pH 6,5-6,0).

Fase 2. Fase de fermentación.

Se inicia al producirse un ambiente anaerobio. Puede durar de días a semanas dependiendo de las características del material ensilado y de las condiciones ambientales en el momento del ensilaje. Si la fermentación se desarrolla con éxito, la actividad BAC (bacterias que producen ácido láctico) proliferará y se convertirá en la población predominante. Debido a la producción de ácido láctico y otros ácidos, el pH bajará a valores entre 3,8 a 5,0.

Las bacterias que producen ácido láctico (BAC) pertenecen a la microflora epifítica de los vegetales. Los componentes BAC que se asocian con el proceso de ensilaje pertenecen a los géneros: Lactobacillus, Pediococcus, Leuconostoc, Enterococcus, Lactococcus y Streptococcus. La mayoría de ellos son mesófilos, o sea que pueden crecer en un rango de temperaturas que oscila entre 5° y 50 °C, con un óptimo entre 25° y 40 °C. Son capaces de bajar el pH del ensilaje a valores entre 4 y 5, dependiendo de las especies y del tipo de forraje.

Fase 3. Fase estable.

La mayoría de los microorganismos de la fase 2 lentamente reducen su presencia. Algunos microorganismos acidófilos sobreviven este período en estado inactivo; otros, como clostridios y bacilos, sobreviven como esporas. Sólo algunas proteasas y carbohidrasas, y microorganismos especializados, como Lactobacillus buchneri que toleran ambientes ácidos, continúan activos pero a menor ritmo. Si el ambiente se mantiene sin aire ocurren pocos cambios.

Fase 4. Fase de deterioro aerobio.

Ocurre en todos los ensilajes al ser abiertos y expuestos al aire para su empleo, pero puede ocurrir antes por daño de la cobertura del silo (p. ej. roedores o pájaros). El período de deterioro puede dividirse en dos etapas. La primera se debe al inicio de la degradación de los ácidos

orgánicos que conservan el ensilaje por acción de levaduras y ocasionalmente por bacterias que producen ácido acético. Esto aumenta el valor del pH, lo que permite el inicio de la segunda etapa de deterioro; en ella se constata un aumento de la temperatura y la actividad de microorganismos que deterioran el ensilaje, los bacilos. La última etapa también incluye la actividad de otros microorganismos aerobios, también facultativos, como mohos y enterobacterias.

1.2.4 CALIDAD DE LOS ENSILADOS.

La calidad de un ensilado depende por una parte de su valor nutritivo, que está ligado directamente a su composición química (fibra, materias nitrogenadas, minerales, hidratos de carbono, etc.) y por otra de la calidad de conservación, que viene definida por los productos finales de las fermentaciones que tienen lugar en el mismo.

La calidad, también puede ser evaluada visualmente, pues aunque esta información no es precisa, en conjunción con el análisis químico, factores como el olor, color o apariencia general puede suministrar una buena indicación del valor nutritivo esperado.

El análisis químico del ensilado resulta complejo ya que su valor nutritivo no solamente viene dado por su contenido en principios nutritivos sino que hay que evaluar si la fermentación ha sido correcta y si, por tanto, el ensilado va a ser estable.

Los ensilados se pueden diferenciar principalmente en tres indicadores de calidad:

Indicadores de calidad nutritiva:

Para poder valorar un ensilado desde el punto de vista nutricional los parámetros a evaluar son pH, materia seca, cenizas, proteína bruta, fibra neutro detergente y estimación de la digestibilidad de la materia orgánica.

La materia seca.- indica que cuando el contenido de MS en el material a ensilar sobrepasa el 25%, se reduce el nivel de efluentes y las pérdidas de carbohidratos por esta vía; McDonald (1981), además disminuye las pérdidas por respiración, que permite un predominio de las bacterias ácido-lácticas y un pH adecuado. Su valor óptimo para la conservación se sitúa entre 30 y 35% (Ojeda et al., 1991).

El pH.- Es un indicador de vital relevancia en el proceso de conservación de un forraje en forma de ensilaje debido a que es una de las transformaciones más radicales que ocurren en el forraje y por su estrecha relación con los procesos degradativos durante la conservación. Es necesario que el descenso del pH ocurra lo más pronto posible para garantizar un hábitat desfavorable para las bacterias clostrídicas y reducir la respiración. El valor de pH está en función de la materia seca del ensilaje y de la proporción que exista entre las proteínas y los carbohidratos solubles, se considera que cuando un ensilaje alcanza valores inferiores a 4.2 se ha logrado su estabilidad fermentativa.

Las cenizas.- nos indican el contenido mineral. Si el porcentaje es alto, mayor del 15 % sobre MS, indica que hubo contaminación con tierra.

La proteína bruta.- (PB) es un parámetro importante debido a su influencia directa en la producción animal. Para ensilados de maíz planta entera, el contenido ha de estar comprendido entre 8 y 10 % sobre MS.

La fibra neutro detergente.- (FND) es la fracción del forraje que se corresponde a las paredes celulares y, está asociada, negativamente con la ingestión de materia seca. El porcentaje de FND se incrementa con el estado de madurez de los forrajes.

- **Indicadores de calidad fermentativa:**

Nitrógeno amoniacal como porcentaje del nitrógeno total.- Para poder utilizarlo en los ensilajes bien conservados se considera como óptima una concentración menor de 7% de nitrógeno amoniacal como porcentaje del nitrógeno total (Ojeda et al., 1991).

Ácidos grasos volátiles.- Dentro de los ácidos orgánicos formados durante la fermentación, el más importante es el ácido láctico, por la alta acidez que induce en el medio y además por ser el resultado del metabolismo de las bacterias más eficientes y adaptadas entre todas las presentes en los ensilajes, lo que permite cumplir una acción bactericida, conservando mejor el ensilaje (Ojeda et al., 1991). Los ácidos acético y butírico deben estar ausentes o en cantidades despreciables. Son el resultado de fermentaciones inducidas por la presencia de bacterias coliformes que transforman el láctico en acético y de gérmenes butíricos, presentes en el estiércol, tierra y especies adventicias que degradan el nitrógeno proteico del forraje en fermentación.

Azúcares solubles residuales.- cuya escasez en el jugo es indicativa de una correcta fermentación láctica.

Un ensilado se puede considerar bien fermentado cuando presenta las siguientes características, aunque hay diferentes baremos según países (INRA, 1981):

- Alto contenido en ácido láctico: Superior al 1,5-3 %, referido a MS.
- Contenido medio en ácido acético: 0,5-1 %, referido a MS.
- Ausencia o trazas de ácido butírico.
- Nitrógeno amoniacal respecto al nitrógeno total: Inferior al 5-10 %.
- Nitrógeno soluble respecto al nitrógeno total: Inferior al 50 %.

• Indicadores organolépticos

Este indicador está basado en la apreciación subjetiva de la calidad de un ensilaje a través de los sentidos y sujeta a la experiencia del personal evaluador. Los parámetros a considerar son: olor, color, textura y grado de humedad. Los rangos de evaluación de acuerdo con las características organolépticas son los siguientes:

Excelente calidad

- **Color:** verde aceituna
- **Olor:** agradable, de fruta madura.
- **Textura:** el forraje conserva todos sus contornos definidos, se aprecian sus vellosidades si las tenía el forraje original, las hojas permanecen unidas a los tallos.
- **Humedad:** no humedece las manos al ser comprimido dentro del puño, con una presión normal se mantiene suelto el ensilaje.

Buena calidad

- **Color:** verde amarillento, los tallos con tonalidad más pálida que las hojas.
- **Olor:** agradable, ligero olor a vinagre. No deja residuos en las manos al ser tocado.
- **Textura:** el forraje conserva todos sus contornos definidos, se aprecian sus vellosidades si las tenía el forraje original, las hojas permanecen unidas a los tallos.
- **Humedad:** No humedece las manos al ser comprimido dentro del puño, con una presión normal se mantiene suelto el ensilaje.

Regular calidad

- **Color:** verde oscuro. Tallos y hojas con igual tonalidad.
- **Olor:** ácido, con fuerte olor a vinagre. Deja en las manos un permanente olor a manteca rancia característico de ácido butírico.

- **Textura:** las hojas se separan fácilmente de los tallos; los bordes del forraje aparecen mal definidos; las hojas tienden a ser transparentes; muy amarillos los vasos leñosos.
- **Humedad:** al ser comprimido en el puño gotean efluentes, con tendencia a ser compactado y formar una masa.

Mala calidad

- **Color:** casi negro o negro.
- **Olor:** desagradable, con olor putrefacto a humedad. Deja un olor a manteca rancia en las manos, el cual permanece por horas.
- **Textura:** no se aprecia diferencia entre hojas y tallos, los cuales forman una masa amorfa, jabonosa al tacto.
- **Humedad:** destila líquido efluente, se compacta con facilidad y llega a tomar la forma deseada.

1.3 APLICACIÓN DE LA TECNOLOGÍA NIRS EN EL ANALISIS DE SUBPRODUCTOS Y ENSILADOS.

El empleo de la espectrometría NIR para la determinación no destructiva de parámetros internos de calidad en alimentos está ampliamente extendida desde los años 60, como técnica para su uso en alimentos.

Los alimentos representan muestras complejas que contienen agua, proteínas, grasas y carbohidratos como componentes mayoritarios. Por ello, la introducción de la reflectancia en el infrarrojo cercano (NIRS) para el estudio y valoración de alimentos supuso un gran avance por su rapidez de respuesta y por ser una tecnología limpia que no emplea reactivos ni genera residuo alguno.

La calidad de un ensilado resulta de la interacción entre la naturaleza del material de origen, su composición química y el proceso de conservación, y viene dada por su contenido en principios nutritivos, la digestibilidad de los mismos y por su fermentación. Su alto potencial

de muestreo y su bajo coste de análisis, contribuyen a la implantación de los procesos de análisis y control.

La técnica NIRS ha contribuido a la valoración rápida de los alimentos para el ganado permitiendo tomar decisiones rápidas, estratégicas y económicas que permiten decidir el tipo de suplementación alimenticia más adecuada en cada momento (Deaville y Flinn, 2000).

1.3.1 ESPECTROCOPIA NIR.

1.3.2 INTRODUCCIÓN A LA ESPECTROCOPIA NIR.

El descubrimiento de la luz infrarroja se debe a Frederick William Herschel (1738- 1822), quien descubrió el denominado espectro al hacer incidir la luz solar a través de un prisma, que la dispersaba en un arco iris. El espectro contiene todos los colores visibles que componen la luz del sol. El proceso que siguió para el descubrimiento fue el siguiente: Herschel hizo pasar luz solar a través de un prisma de cristal para generar el espectro: el arco iris, el cual se forma cuando la luz se divide en los colores que la componen. Luego midió la temperatura de cada color. Para ello Herschel utilizó tres termómetros con bulbos ennegrecidos para absorber el calor. Colocó un bulbo en cada color, mientras que otros dos fueron colocados fuera del espectro, como muestras de control. Al medir las temperaturas de la luz violeta, azul, verde, amarilla, naranja y roja notó que cada color tenía una temperatura mayor que los termómetros de control, y que la temperatura de los colores del espectro aumentaba al ir del violeta al rojo. Al medir la temperatura en la zona cercana a la luz roja del espectro, desprovista de la luz, vio que la temperatura era la más alta de todas. A esta radiación invisible la denominó radiación infrarroja.

Pero hasta la década de los 50, con la aparición de los primeros espectrofotómetros NIR, fue cuando se desarrollaron las primeras aplicaciones de esta región del espectro electromagnético.

En los años 60, el grupo de investigación del Departamento de Agricultura de EEUU (USDA) dirigido por Karl Norris, utilizó la espectroscopía NIR para el estudio de muestras complejas de origen agroalimentario (Norris, 1964; Ben-Gera y Norris, 1968). Demostrando el gran potencial analítico de esta técnica, de gran aplicabilidad en multitud de sectores.

La nueva instrumentación con ópticas mejoradas, a partir de finales de los años 70 se hizo posible el registro de espectros de manera rápida y reproducible (Henry, 1999). Sin embargo, el auge de la espectroscopía NIR como técnica de análisis se ha dado en los años 80 y 90, impulsada por los grandes avances tecnológicos y la expansión en el sector de la informática.

A diferencia de otras técnicas espectroscópicas, en el infrarrojo cercano no se generan picos aislados característicos de cada compuesto, sino que se da una superposición de bandas correspondientes a diferentes grupos funcionales.

Esta tecnología en la actualidad, se aplica en sectores como los agroalimentarios, la industria química, petroquímica y textil, medicina, farmacia, cosmética, biología y medio ambiente. En las industrias agrarias y alimentarias, la tecnología NIRS se ha venido utilizando desde hace tiempo para el análisis y control de numerosos productos tales como piensos y cereales, lácteos, aceites y grasas, productos cárnicos, frutas, vegetales, etc. (Davies y Grant, 1987; Nicolai *et al.*, 2007; Cen y He, 2007).

1.3.3 VENTAJAS E INCONVENIENTES DE LA TECNOLOGÍA NIRS.

Como ventajas directas de la espectroscopía NIR, por lo que se usa actualmente, podemos destacar las siguientes (Eliécer *et al.*, 2001):

- Es una técnica no destructiva.
- El tiempo requerido para obtener el resultado analítico de una muestra escaneada en el equipo es inferior a 1 minuto.

- La muestra no requiere preparación.
- No se requieren reactivos químicos para efectuar el análisis.
- Es una técnica limpia que contribuye a la conservación del medio ambiente.
- Somos capaces de repetir la medida varias veces para una misma muestra.
- La operación del equipo es sencilla.
- Se pueden analizar múltiples constituyentes simultáneamente en la misma muestra.
- Es una técnica muy versátil: si las muestras contienen enlaces tales como C-H, N-H, O-H, o si la concentración de lo analizado excede alrededor de 0,1% de la composición total, entonces es muy posible llegar a respuestas aceptables.
- Es capaz de predecir características químicas, físicas y sensoriales, ideal para ser implementada en sistemas de aseguramiento de la calidad (Garrido, 1997).
- Podemos cuantificar una sustancia en una mezcla compleja, sin hacerse necesaria una separación previa (Skoog y Leary, 1996).

En cuanto al tiempo requerido para obtener resultados, el equipo de tecnología NIRS posee una alta velocidad de respuesta. Un ejemplo de esto es que hoy día un técnico es capaz de analizar entre 15 y 20 muestras para obtener un dato de un constituyente como la grasa, en ocho horas de trabajo. En cambio, con la tecnología NIRS de “tercera generación” somos capaces de analizar 600 espectros por segundo.

Al no usarse productos químicos se respeta el medio ambiente. Un ejemplo de esto es que durante el curso internacional “Applications of NIRS technology for the evaluation of agriculture products”, desarrollado en la ETSIA de Córdoba en 1994, se presentaron algunas estimaciones

de los residuos producidos en determinados análisis. Así, un laboratorio que realizara 5.060 análisis Kjeldahl al año durante 15 años, produciría 2,76 t de hidróxido sódico y 900 t de agua de refrigeración. (Garrido, 1997).

Por tanto, se debería considerar la reducción en contaminación ambiental que podría obtenerse mediante el uso del NIRS en análisis de rutina (Garrido, 1997).

El trabajo preliminar es el típico de cualquier método quimiométricos. Aún después de varias horas o días de calibración, las múltiples ventajas del análisis por NIRS son más importantes que el tiempo requerido para el desarrollo del método.

Sin embargo el método también tiene desventajas como las siguientes (Garrido, 1999):

- Utilizada principalmente como técnica indirecta.
- Influenciada por factores tales como la temperatura, la humedad y el tamaño de partícula.
- Alto coste de inversión inicial.
- El proceso de calibración y validación demanda tiempo.
- Se requiere una formación multidisciplinar: espectroscopía, estadística, matemáticas, química, tecnología de los alimentos, etc.
- Altamente dependiente del desarrollo en instrumentación, software, quimiometría, etc.

1.3.4 BASES TEORICAS DE LA ABSORCIÓN NIR.

La región del infrarrojo (IR) es aquella región del espectro electromagnético comprendida entre el final de la región visible y la región de microondas. Dentro de dicha región, se pueden diferenciar distintas secciones en función del intervalo de longitudes de onda

considerado, aunque las fronteras entre unas y otras son un tanto arbitrarias.

Los datos de una muestra que se obtienen mediante un espectrómetro NIR contienen información espectral relacionada con la composición de las muestras estudiadas. Las diferentes relaciones químicas que componen la muestra tienen absorción a unas determinadas longitudes de onda lo que se puede emplear para definir la composición química de diferentes sustancias. La obtención de medidas de esta manera resultaría fácil. Pero existen problemas que dificultan su medida como que los picos de absorción de las diferentes moléculas químicas se solapan en muchas partes de la región espectral. La señal de absorción es a menudo débil si se compara con otros fenómenos que intervienen en la interacción entre la luz y partícula, como por ejemplo, todos los efectos de dispersión. Por estas razones, es necesario procesar matemáticamente los datos espectrales a fin de extraer información valiosa de propiedades químicas de las muestras (Arruabarrena, 2012). La forma de cuantificar la absorción de energía en la región NIR es a través de la medida de la energía reflejada o transmitida. La 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 Ciuczak, 1992; Osborne *et al.*, 1993; ASTM, 1994; Bertrand y Dufour, 2000).

La tecnología NIRS se basa en la ley de Lambert-Beer, según la cual la 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, 1987ab; Burns y Ciuczak, 1992; Osborne *et al.*, 1993; Shenk y Westerhaus, 1995a; Bertrand y Dufour, 2000). Por ello, los valores de absorbancia obtenidos a partir de una muestra se pueden relacionar con la concentración de

determinadas sustancias presentes en la muestra, siempre y cuando absorban radiación NIR.

Cuando un compuesto orgánico es sometido a radiación infrarroja los grupos funcionales (-OH, -NH, -CO y -CH) sufren una serie de vibraciones y proporcionan bandas características a determinada longitud de onda. La energía que se absorbe, se recoge en forma de espectros que son la representación gráfica de esta absorción.

Para llevar a cabo el análisis cuantitativo, es necesario tener un conjunto de espectros con sus respectivos datos de laboratorio y relacionarlos mediante un tratamiento matemático y estadístico para obtener una ecuación de calibración. Cuando se tiene una muestra cuya composición se quiere determinar, ésta se escanea para obtener el espectro y mediante la ecuación de calibración se obtiene el dato deseado. El NIR permite hacer calibraciones para muchos productos y componentes, siempre que se disponga de análisis físico químico de referencia, preciso.

1.3.5 ESPECTRO ELECTROMAGNETICO.

El espectro que interesa (NIR), está comprendido entre las longitudes de onda 400-2500, siendo el visible el que está entre las 400-700 nm y el espectro infrarrojo el que se encuentra situado entre las 780-2500 nm. **(Tabla 2).**

Tabla 2. Tipos de espectros.

ESPECTRO ELECTROMAGNÉTICO		
Tipo de espectro	Intervalos de longitudes (nm)	
Rayo X	$10^2 - 10$	
Ultra violeta	10 - 100	
Espectro Visible	Violeta	400 - 435
	Azul	435 - 475
	Verde	475 - 512
	Amarillo	512 - 580
	Naranja	580 - 600
	Rojo	600 - 670
NIR	VIS-NIR	400 - 2500
	NIR (Reflectancia)	1100 - 2500
	NIR (Transmitancia)	800 - 1380
IR	1000 - 10000	
Ondas de Radio	$10000 - 10^{16}$	

Fuente: Skoog y Leary, 1996.

La región infrarroja del espectro incluye la radiación con longitudes de onda comprendidas entre los 780 y los 106 mm. Tanto desde el punto de vista de las aplicaciones como de los instrumentos, es conveniente dividir el espectro infrarrojo en tres regiones denominadas infrarrojo cercano, medio y lejano. (Tabla 3).

Tabla 3. Espectro de longitudes de onda

REGIÓN	INTERVALO DE LONGITUD DE ONDA (λ), NM	INTERVALO DE NÚMERO DE ONDA ($\tilde{\nu}$) CM^{-1}	INTERVALO DE FRECUENCIAS (ν), HZ
Cercano	780-2500	12800-4000	$3,8 \times 10^{14} - 1,2 \times 10^{14}$
Medio	2500-5000	4000-400	$1,2 \times 10^{14} - 6,0 \times 10^{12}$
Lejano	$5000 - 10^6$	400-10	$6,0 \times 10^{12} - 3,0 \times 10^{11}$

Fuente: Skoog y Leary, 1996

1.3.6 QUIMIOMETRÍA Y SU APLICACIÓN EN LA TECNOLOGÍA NIRS.

La quimiometría es la parte de la química que, usando métodos matemáticos, estadísticos y de lógica formal diseña o selecciona procedimientos de medida óptimos y proporciona la máxima información

relevante (Massart *et al.* 1988). La utilización de métodos quimiométricos permite la identificación de muestras, la posibilidad de determinar simultáneamente distintos analitos, Como ventaja fundamental de la quimiometría podríamos destacar el conocimiento más amplio del problema y una alta velocidad de análisis, lo que permite reducir costes y tiempo de análisis.

El uso de la tecnología NIRS, se debe fundamentalmente a que en la actualidad la determinación de los parámetros de calidad, se realizan por procedimientos, en la mayoría subjetivos (color y textura), o la medida de la fecha de madurez (contenidos en sólidos solubles y ácidos). Por ello, existe una fuerte demanda de técnicas objetivas, rápidas y no destructivas. Estudios recientes han aportado rapidez y comodidad en la medida de muestras.

La tecnología NIRS, reduce considerablemente los costes por muestra en comparación con los métodos químicos. Esta técnica puede reemplazar a otros métodos tradicionales como la refractometría. Además no precisa una preparación previa de la muestra y no usa ningún tipo de reactivo químico por lo que no genera residuos y puede considerarse una técnica más sostenible. También elimina las fuentes de error comunes en la manipulación de muestras en el laboratorio. En la práctica, los análisis NIRS, son mucho más versátiles y más rápidos que otras técnicas espectroscópicas (Jarén *et al.*, 2001).

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.

La extracción de información relevante de los espectros NIR se complica debido a la existencia de múltiples fuentes de error que afectan

al dato espectral. Dichas fuentes de error se pueden relacionar con cada uno de los siguientes aspectos:

- relacionadas con el instrumento (relación señal/ruido, exactitud, y precisión de la longitud de onda, linealidad de la señal, tipo y tamaño de la cápsula, etc.)
- relacionadas con la muestra (homogeneidad, densidad, textura, granulometría, estabilidad ante fluctuaciones de la temperatura, etc.)
- operacionales (preparación de la muestra, errores del analista, estrategias y procedimientos estadísticos empleados, etc.).

Como consecuencia de todo lo expuesto se hace necesario emplear herramientas quimiométricas complejas en el tratamiento de los datos espectrales con el fin de minimizar estos errores y obtener información relevante (Naes *et al.*, 2002). Entre tales herramientas quimiométricas cabe destacar los métodos de análisis multivariante.

El análisis multivariante podría definirse como los diferentes métodos estadísticos, matemáticos o gráficos, empleados en el análisis de datos, siempre y cuando se consideren varias variables de forma simultánea, (Martens y Martens, 2000, Naes *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).

1.3.7 ETAPAS DEL DESARROLLO DE MODELOS DE CALIBRACIÓN MULTIVARIABLES.

Un modelo de calibración multivariable relaciona las N variables experimentales (por ejemplo, un espectro NIR) frente a propiedades conocidas de las muestras (por ejemplo, la concentración de un analito), siendo el objetivo la predicción de dicha propiedad en nuevas muestras.

1.3.8 OBTENCIÓN DE LOS ESPECTROS DE ABSORCIÓN.

La técnica NIR comienza con la obtención de un espectro de absorción (entre 750 y 2.500 nm) a partir del escaneado de la muestra. El espectro refleja las transiciones cuantizadas vibracionales producidas en las moléculas como consecuencia de la interacción entre la radiación emitida por una lámpara halógena de filamento de tungsteno y la muestra.

1.3.9 SELECCIÓN DEL CONJUNTO DE MUESTRAS.

El primer paso es disponer de un número de muestras suficiente para la creación y validación del modelo. Las muestras utilizadas para el desarrollo de un modelo de calibración deben ser representativas de la variabilidad que pueda darse durante el proceso, y también entre procesos. Así, las muestras, además de cubrir todo el intervalo de la propiedad a determinar, han de incorporar la variabilidad debida a otras variables del proceso, como puede ser el origen de la materia prima o la temperatura de reacción, entre otros. Una herramienta útil para la selección de muestras es el análisis en componentes principales o PCA.

1.3.9.1 MÉTODO DE REFERENCIA.

El desarrollo de modelos cuantitativos requiere el conocimiento previo de las variables que van a ser determinadas. Para ello, se utilizan los métodos de referencia (análisis por vía húmeda) que deben proveer valores precisos y exactos, ya que de ello dependerá la calidad del modelo a desarrollar.

1.3.9.2 PRETRATAMIENTO ESPECTRAL.

Los pretratamientos espectrales tienen como objetivo, entre otros, minimizar estas contribuciones espectrales no deseadas, simplificando los modelos y con esto, acrecentar la señal del analito de interés. No todos los pretratamientos consiguen estos objetivos simultáneamente, por lo que la selección del más apropiado es puramente experimental.

Se entiende como pretratamientos independientes del grupo de muestras aquellos que tratan cada espectro de forma individual sin tener en cuenta el resto de muestras o datos de concentración; y como dependientes aquellos que tratan los espectros en conjunto o que requieren datos de concentraciones de referencia, por lo que el pretratamiento varía según el conjunto de datos. A pesar de que los pretratamientos más utilizados han sido los independientes del grupo de muestras, cabe destacar la importancia de los métodos dependientes para el estudio de los datos espectrales

Los pretratamientos espectrales más utilizados son:

1. Pretratamientos independientes del grupo de muestras.

Suavizado Espectral: Los métodos de suavizado son utilizados para reducir matemáticamente el ruido aleatorio o incrementar la relación señal/ruido (Beebe *et al.*, 1998). Estos métodos utilizan una ventana de puntos para determinar un valor central, desplazando la ventana a lo largo de todo el espectro. Los métodos más utilizados son el de Savitzky-Golay (Savitzky y Golay, 1964) y el de la Transformada de Fourier (Horlick, 1972).

El SNV o *Standard Normal Variate*: se utiliza principalmente para corregir variaciones de la línea base producidas por la dispersión, (Barnes *et al.*, 1989) por lo que puede ser útil para sólidos o muestras líquidas con turbidez. Este pretratamiento se centra cada espectro, restando el valor medio de la absorbancia del espectro a los valores de

absorbancia de cada longitud de onda, dividido todo ello por la desviación estándar:

$$x'_n = \frac{x_n - \bar{x}}{s}$$

Derivadas: Es uno de los pretratamientos más utilizados en espectroscopía NIR por su capacidad de solventar los problemas más característicos de esta técnica: el solapamiento de bandas y los desplazamientos de la línea base. La primera derivada elimina los desplazamientos de la línea base constantes y la segunda derivada elimina los desplazamientos que varían linealmente con la longitud de onda. Los métodos más utilizados son el de Savitzky-Golay (Savitzky y Golay, 1964) y el de Norris (Norris y Williams, 1984). Las derivadas por el método de Savitzky-Golay incorporan un suavizado, por lo que el tamaño de ventana es un parámetro crítico. Si éste es pequeño la derivada tendrá elevado ruido y si es grande la información espectral resultará demasiado suavizada (Beebe *et al.*, 1998).

2. Pretratamientos dependientes del grupo de muestras.

Promedio de Espectros: El ruido es considerado una variación instrumental aleatoria, por lo que promediando varias señales analíticas de la misma muestra, el ruido disminuye y se mejora la relación señal/ruido.

Centrado: El centrado es un pretratamiento habitual previo al desarrollo de un modelo de calibración o de un análisis de componentes principales. Consiste en centrar cada una de las N variables restando la media de todos los elementos de esa variable. Los vectores centrados tienen valor medio igual a cero, sin modificar la varianza. Un aspecto que influye de manera importante en la baja selectividad de la información espectral NIR es la dispersión de la radiación incidente, también denominada **efecto scatter** (Bertrand y Dufour, 2000). Este hecho es

debido 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, lo que provoca numerosas variaciones no deseadas en los datos espectrales NIR (Barnes *et al.*, 1989; Dhanoa *et al.*, 1994; Fernández y Garrido, 1999; Naes *et al.*, 2002).

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

El MSC o *Multiplicative Scatter Correction*: se utiliza para compensar efectos multiplicativos o aditivos debidos al *scattering*. El MSC corrige los desplazamientos entre muestras aplicando a los espectros originales los parámetros de la recta, obtenida por mínimos cuadrados, representando los valores de absorbancia de un intervalo de longitudes de onda en el que no absorba el analito (aunque a la práctica se realiza con el espectro completo), frente a los valores de absorbancia de un espectro de referencia (generalmente el espectro medio de los datos a tratar). La corrección del espectro original se reduce a la siguiente fórmula, en la que a y b son los parámetros de la recta.

$$x_i^{MSC} = \frac{x_i - a}{b}$$

El DT o *detrending*: es un tratamiento para quitar los efectos de *offset* en la línea de base, así como la pendiente o curvatura de los espectros, normalmente causados por el *scatter*. El cálculo de esta función se resta del espectro original, dando como resultado el espectro

con tratamiento *detrend* aplicado. Normalmente se aplica conjuntamente a otros tratamientos (por lo general el SNV).

1.3.10 DESARROLLO DEL MODELO DE CALIBRACIÓN

Seleccionadas las muestras de calibración, y analizadas éstas mediante el método de referencia y NIRS, se busca una relación entre los datos espectrales y analíticos. Para esta finalidad se utilizan diferentes tratamientos estadísticos como MLR; PLSR; PCA.

El procedimiento *MLR (REGRESION LINEAL MULTIPLE)*: Este algoritmo calcula el modelo de regresión en base a todas las variables espectrales disponibles y los valores analíticos, por lo que no se realiza una simplificación de las variables (Shenk y Westerhaus, 1995b; Puigdomenec, 1998). Las longitudes de onda seleccionadas, mediante un test F, serán aquellas que presentan una mayor correlación entre el valor de absorbancia a una longitud de onda determinada y el dato de composición química.

El modelo estadístico PCA (COMPONENTES PRINCIPALES), se basa en la descomposición de la matriz de datos espectrales mediante un PCA, por lo que genera unas nuevas variables ortogonales no correlacionadas, con las cuales desarrolla una regresión lineal junto con la matriz de datos que contiene la información de referencia. El inconveniente de este algoritmo es que sólo usa la información espectral para calcular las nuevas variables. Otro modelo ampliamente utilizado es *PLSR (mínimos cuadrados)*, es el utilizado en este trabajo. Este tratamiento se diferencia del anterior en que las nuevas variables ortogonales no sólo se calculan a partir de la matriz de datos espectrales, sino que considera también la matriz de los datos de referencia. De este modo, las nuevas variables maximizan la varianza explicada de la matriz espectral (X) de forma que sea relevante para predecir los parámetros de la matriz con los datos de referencia (Y) (Martens y Naes, 1989; Wold *et al.*, 2001). Los métodos de regresión PCR y PLSR se asemejan en sus fundamentos. Estos métodos de regresión permiten

obtener una relación lineal entre los datos espectrales y los de su 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 (Martens y Naes 1989; Shenk y Westerhaus, 1996; Bertrand y Dufour, 2000; Naes *et al.*, 2002). Pero existen muestras y espectros anómalos que afectan a las ecuaciones NIRS obtenidas por los procedimientos estadísticos mencionados anteriormente y a su capacidad predictiva (Williams, 1987). Esos datos anómalos llamados **outliers (Figura 1)** se consideran datos erróneos, inusuales o atípicos comparados con el resto de datos de las muestras de calibración. Las causas de esa anomalía puede ser por el error en el hardware, error del operador, errores en los datos de referencia, errores en los datos espectrales, muestras y espectros poco representados dentro del grupo de calibración. Con esos datos es importante no eliminarlos hasta no tener una explicación clara, de ahí que para tomar esa decisión se tiene que tener en cuenta unos parámetros estadísticos como la T ("t de Student), H (distancia de Mahalanobis), entre otros.

La T-test de cada longitud de onda, señala la importancia de cada longitud dentro de la calibración, la que presente mayor valor será la más importante para la calibración. Si el valor t es superior a 10 se considera esencial para formar parte de la ecuación de calibración.

Ecuación test-T

$$t = \frac{Y_{pred} - Y_{ref}}{ETC * \sqrt{1 - H}}$$

ETC: Error típico de calibración
(outliers)

H: Error del espectro

En análisis NIRS para $t > 2.5$ se consideran significativos.

Para la detección de anómalos espectrales, entre los métodos más utilizados se encuentra el cálculo de distancias n-dimensionales, como puede ser la distancia de Mahalanobis (se consideran como pertenecientes a la población aquellas muestras con $GH < 3$) y el cálculo de residuales en los datos espectrales.

-Para el modelo MLR se expresa con la siguiente ecuación.

$$H = \frac{K}{n}$$

-Para los modelos PLSR y PCR se expresa mediante la ecuación

$$H = \frac{K+1}{n}$$

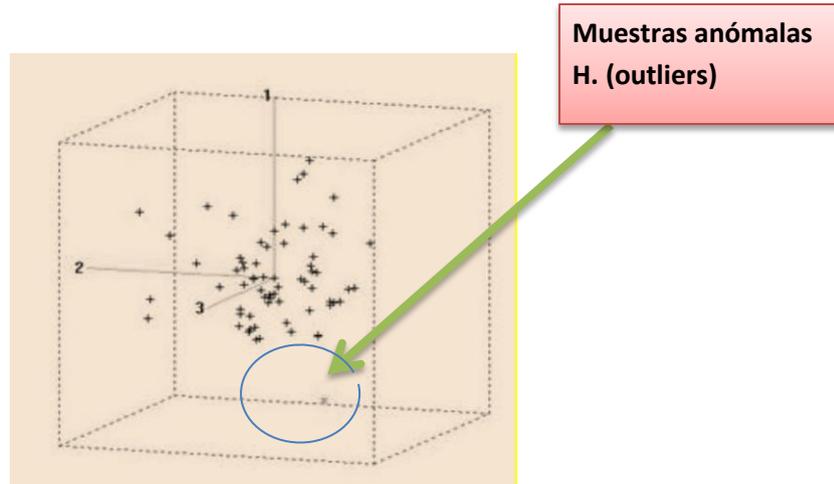
H: Distancia de Mahalanobis.

n: nº de espectros en el conjunto de la calibración.

K: nº de longitudes de onda seleccionadas.

Ambos tipos de estadísticos (distancias en espacios n-dimensionales y residuales) pueden ser también empleados en la detección de anómalos químicos. El cálculo de residuales está más extendido en lo que a datos de composición se refiere. Dichos residuales químicos serían aquellas muestras que presentan diferencias significativas entre los valores de composición aportados por el método de referencia y los estimados mediante el modelo de regresión, evaluados normalmente mediante un test T

Figura 1. Representación tridimensional de la población de muestras.



1.3.11 EVALUACIÓN DEL MODELO DE CALIBRACIÓN.

El modelo de calibración se evalúa inicialmente mediante un proceso de validación cruzada, en el que se calculan distintas ecuaciones y se procede a seleccionar aquella que mejor predice las muestras.

Stone (1974) describe esta técnica que 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.

1.3.12 PARÁMETROS PARA EVALUAR EL MODELO DE CALIBRACIÓN.

Después de desarrollar distintas ecuaciones de calibración mediante alguno de los métodos de regresión existentes se ha de validar nuevamente la fiabilidad del modelo de calibración mediante la

predicción de muestras que no hayan intervenido en la misma, utilizando para ello diversos estadísticos. (Etapa de validación).

La etapa de validación consiste en la comparación del valor verdadero y el valor estimado (Massart *et al.*, 1988). Considerando, en este caso, el valor verdadero a aquel obtenido mediante el método de referencia, mientras que el valor estimado será el predicho por la ecuación NIRS.

Los estadísticos empleados en la evaluación, selección y validación de las ecuaciones de calibración son:

- **(Coeficiente de determinación de la calibración (R^2) / Coeficiente de determinación de la validación cruzada (r^2)).** Este parámetro expresa la bondad de ajuste del modelo de las ecuaciones de calibración. Establece una correlación entre los datos analíticos obtenidos en el laboratorio y los predichos por las ecuaciones de calibración para cada uno de los componentes analizados. Se busca el valor de más próximo a 1. Según Shenk y Westerhaus (1996), establecen unos intervalos de R^2 para establecer la precisión de las ecuaciones obtenidas. Dichos intervalos se pueden apreciar en el siguiente cuadro:

- Si $R^2 > 0,90$ excelente precisión.
- Si $0,89 > R^2 > 0,70$ buena precisión
- Si $0,69 > R^2 > 0,50$, buena separación entre valores bajos, medios y altos.
- Si $0,30 > R^2 > 0,49$ correcta separación entre valores bajos y altos.
- Si $0,05 > R^2 > 0,29$ mejor no analizar.

-RMSEE (media de la raíz cuadrada del error de calibración/estimación): Es el error asociado a las diferencias existentes entre los análisis analíticos determinados en el laboratorio por los métodos de referencia y los resultantes del análisis mediante la tecnología NIRS para cada uno de los parámetros determinados en las muestras utilizadas en la calibración. Este parámetro estadístico se requiere sea el menor posible.

$$RMSEE = \sqrt{\frac{1}{M-R-1}} SSE$$

M: nº de espectros de calibración.

R: nº de factores (rango) empleados (PLSR).

SSE: error estándar de calibración/estimación

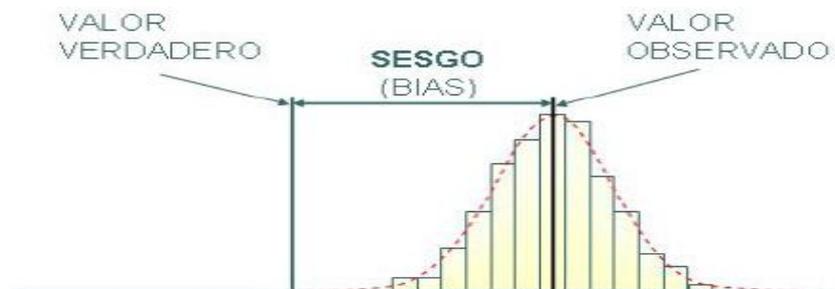
-RMSECV (Error típico de validación cruzada) o error de los residuales para el colectivo de validación cruzada (**RMSEVC**), el cual ha de ser el menor posible. El valor de (**RMSEVC**) final se calcula como la raíz cuadrada de la media de los valores de los (**RMSEVC**) de los distintos pases de validación cruzada al cuadrado.

-RPD. Se define como el cociente entre la desviación estándar de los datos de referencia y el RMSEE/RMSECV. Este cociente según Williams y Sobering (1996). Este cociente ha de ser menor o igual a 3, aunque un valor de 2,5 es aceptable (Williams, 2003). Valores superiores a 3 indicaría anómalos espectrales, es decir, muestra cuyo espectro difiere del espectro medio.

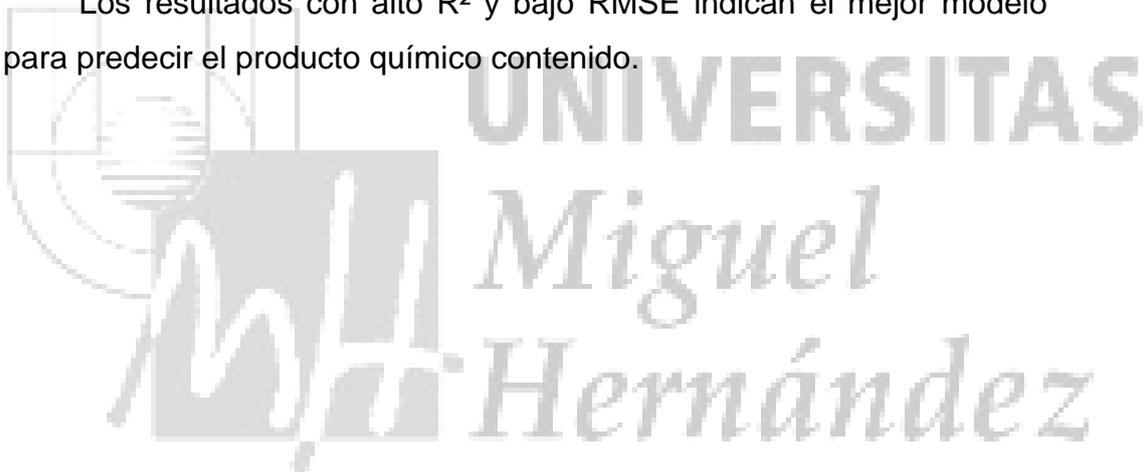
-BIAS. Consiste en la diferencia entre el valor medio predicho por FT-NIRS y el valor medio del modelo de referencia predictivo y la desviación residual (RPD) (Williams, 2001; Williams Y Sobriedad, 1995; Windham, Mertens, y Barton, Donde n es el número de muestras de la calibración, y xi es el resultado que se obtiene por NIRS por otro lado yi es el resultado obtenido por el método de referencia de la i de la muestra.

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)$$

Figura 2. Diferencia entre el valor promedio observado y el valor de referencia



Los resultados con alto R^2 y bajo RMSE indican el mejor modelo para predecir el producto químico contenido.



2. OBJETIVOS

La técnica Espectroscopia en el Infrarrojo Cercano (NIRS) es una técnica rápida, no destructiva y de bajo costo capaz de analizar sustancias orgánicas de forma precisa, fiable y de manera in-situ. Es una técnica ampliamente utilizada en industrias de procesados de alimentos integrando instrumentos en el proceso de elaboración del producto con objeto de poder actuar, de forma instantánea, sobre todos los eslabones de la cadena a fin de garantizar las características del producto final.

Actualmente se han desarrollado aplicaciones con sondas de fibra óptica, que permiten el acercamiento del instrumento a la línea de producción o equipos portátiles que posibilitan el desplazamiento de un laboratorio en miniatura para la realización de medidas in situ, incluso desde el origen (explotaciones agro-ganaderas, a pie de campo ante la recogida de material vegetal etc...). Incluso Comités Internacionales ya han aceptado la técnica NIR como un método de análisis para cuantificar componentes.

Con este fin, El objetivo principal de este trabajo es la determinación mediante la técnica NIRS de componentes como la grasa bruta, proteína bruta, fibra bruta, materia seca, materia orgánica y cenizas en muestras de subproductos vegetales para en un futuro poder realizar determinaciones de estos componentes de forma in situ en explotaciones agro-ganaderas o a pie de campo en la recogida de la cosecha.

Para ello se han obtenido distintas ecuaciones de calibración mediante el algoritmo PLS (mínimos cuadrados parciales) con diferentes pretratamientos espectrales para obtener aquella ecuación de calibración que sea más precisa y fiable en la determinación de los componentes citados anteriormente y sobre muestras de subproductos.

3. MATERIALES Y METODOS.

3.1. MUESTRAS Y ANÁLISIS NIRS.

Se recogen un total de 100 muestras de subproductos vegetales en el municipio de Orihuela procedentes de industrias agroalimentarias y de restos de cosecha, en algunos de ellos se utilizó la técnica de ensilado en microsilos para su mejor conservación, entre estos subproductos destacan:

Pulpa de naranja, planta brócoli, hoja morera, naranja deshidratada subproducto brócoli, brácteas alcachofa, ryegrass, pulpa limón, poda verde pulpa de limón, alfalfa, chumbera, hoja caña seca, planta de boniato, silo alfalfa Sardina, brácteas alcachofa, sorgo fresco, subproducto té, planta romanesco, hoja limón, planta de alcachofa picada, melón con planta.

Una vez recogidas las muestras se trasladaron al laboratorio de investigación del Departamento de Tecnología Agroalimentaria/ Área de Producción Animal del Campus de Orihuela, Universidad Miguel Hernández. Una vez las muestras en el laboratorio se desecaron a 60°C, para posteriormente proceder a su molienda hasta conseguir un tamaño de partícula inferior a 0,2 mm y obtener dos alícuotas.

Una primera alícuota se destinó para determinar los componentes químicos (materia seca, grasa bruta, proteína bruta, fibra bruta, materia orgánica y cenizas) con los métodos oficiales de referencia. La segunda alícuota se trasladó a los Servicios Técnicos de Investigación ubicados en el Campus de Desamparados de la Universidad Miguel Hernández, para obtener los espectros mediante el equipo NIRS Bruker Optics.

3.2. INSTRUMENTO Y ANÁLISIS NIRS.

En el desarrollo del presente trabajo se ha empleado el equipo NIRS, Bruker Optics, (**Figura 3**) equipado con una esfera integradora. Las muestras se analizaron de modo intacto, previa homogeneización

manual y correcta carga de la cápsula empleada (**Figura 4**). Todas las muestras se analizaron por triplicado, obteniendo de cada una de ellas un espectro que a su vez fue el promedio de 64 barridos con una resolución de 16 cm^{-1} . El espectro final para cada muestra es el promedio de tres espectros con 64 barridos. Las medidas se tomaron como valores de absorbancia ($\log 1/R$, donde R es la reflectancia) en un rango comprendido entre 4000-12.500 cm^{-1} (800-2500 nm) a intervalos de 1 nm. Los datos espectrales se recogieron y el posterior análisis quimiométrico con el software “Opus Spectroscopy” versión 6.

Figura 3. Equipo NIRS Bruker Optics



Figura 4. Capsula muestras



3.2. ANÁLISIS QUÍMICOS DE REFERENCIA.

3.2.1. DETERMINACIÓN DE MATERIA SECA.

Fundamento:

El método usado para la determinación de la materia seca fue por desecación en estufa a $105^{\circ}\text{C} \pm 2^{\circ}\text{C}$ (AOAC, 1995) hasta peso constante (**Figura 5**). Con esta determinación analítica se pueden evaporar algunas sustancias volátiles de la materia seca obteniendo un error por defecto y otros componentes pueden oxidarse fijando oxígeno en su molécula y produciendo un aumento de peso en la muestra que implica un error por exceso.

Cálculo:

Conocida la pérdida de peso correspondiente a la cantidad de muestra empleada y aplicando la siguiente fórmula se calcula el porcentaje de materia seca:

$$\text{MS (\%)} = \text{E/D} * 100$$

Dónde:

MS = Materia Seca

D = Peso B - Peso A (peso muestra húmeda)

E = Peso C – Peso A (peso muestra seca).

Figura 5. Esquema de la determinación de materia seca.



3.3.2. DETERMINACIÓN DE CENIZAS.

Fundamento:

La fracción inorgánica de las muestras analizadas, denominándose con el término de cenizas, se determina mediante una calcinación a 550°C (AOAC, 1995), (**Figura 6**). Esta fracción inorgánica está constituida básicamente por óxidos, carbonatos, fosfatos y sulfatos minerales. Algunos componentes orgánicos se descomponen dando lugar a anhídrido carbónico, óxidos de nitrógeno, azufre y agua.

Cálculo:

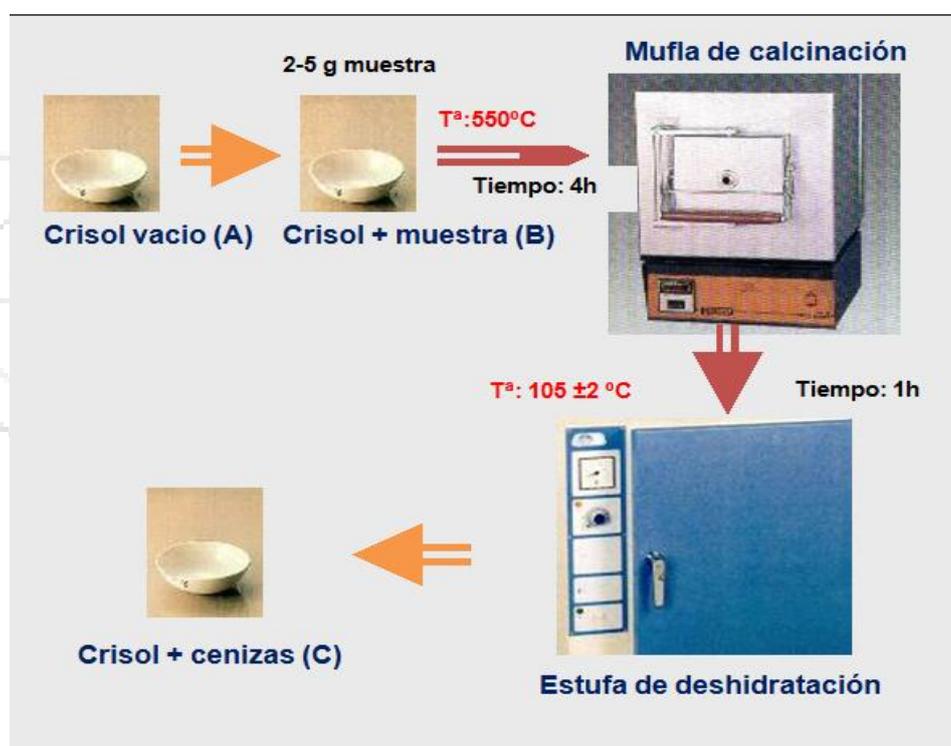
Conocido el peso del residuo incinerado se calcula el contenido de cenizas mediante la siguiente fórmula. El cálculo se realizará referido a materia seca de la muestra.

$$\text{Cenizas (\% MS)} = (E/D * MS) * 100$$

Dónde:

D = Peso B - Peso A (peso muestra húmeda)
E = Peso C – Peso A (peso residuo incinerado)
MS = Materia Seca.

Figura 6. Esquema de la determinación de cenizas



3.3.3. DETERMINACIÓN DE MATERIA ORGÁNICA.

La materia orgánica se calcula con la diferencia entre materia seca y cenizas.

$$\text{Materia Orgánica (\% MS)} = (\% \text{ MS}) - \% \text{ Cenizas (\% MS)}$$

3.3.4. DETERMINACIÓN DE GRASA BRUTA.

Fundamento:

La grasa bruta o extracto etéreo se determinó mediante la metodología AOAC, 1995, la cual, se basa en la solubilidad de los lípidos y otros componentes liposolubles con éter dietílico con previa hidrólisis ácida. El objetivo de la hidrólisis es debido a que existen ciertas muestras que tienen los lípidos ligados a otras sustancias (productos de origen animal, residuos de destilería, gluten, pulpas secas, productos lácteos, piensos enriquecidos con grasa...) o en forma de jabones (heces..) que antes de la extracción con éter deben ser disueltas con ácido clorhídrico. Actualmente se utiliza celite para facilitar más todavía la disolución de jabones. (AOAC, 1995)

Cálculo:

Los cálculos se referirán a la materia seca de la muestra mediante la siguiente fórmula.

$$\text{GRASA BRUTA (\% MS)} = ((P_3 - P_2) / P_1) * 100$$

Dónde:

P_1 = Peso de la muestra en g

P_2 = Peso inicial del vaso

P_3 = Peso final del vaso con la materia grasa

3.3.5. DETERMINACIÓN DE PROTEÍNA BRUTA.

Fundamento:

La determinación de proteína bruta se determinó por el método Kjeldahl (AOAC, 1995). Este método análisis se basa en una mineralización de la muestra con ácido sulfúrico concentrado para transformar todo el nitrógeno en sulfato de amonio. Seguidamente se hace una destilación con NaOH concentrada para transformar el sulfato de amonio en amoniaco. Este amoniaco por arrastre de vapor se recoge sobre ácido bórico y es valorado por titulación con ácido clorhídrico de normalidad 0,1N,

Cálculo:

Los cálculos para la determinación de la proteína bruta son los siguientes:

$$\% \text{ NT} = (1.4 * \text{N} (V_1 - V_0)) / P$$

$$\% \text{ PB (\% MS)} = \text{N (\%)} * 6.25$$

Dónde:

P = Peso en g de la muestra

V₁ = Volumen, en ml de ácido clorhídrico utilizado en la determinación

V₀ = Volumen en ml de ácido clorhídrico utilizado en el ensayo del blanco

3.3.6. DETERMINACIÓN DE FIBRA BRUTA.

Fundamento:

La determinación de la fibra bruta se determinó mediante la metodología descrita por la AOAC, 1995). Es el residuo que queda después de someter la muestra previamente desengrasada y seca a un tratamiento con ácido y álcali diluidos, descontando el valor de las cenizas contenidas en dicho residuo.

Cálculo:

$$\text{FIBRA BRUTA (\% MS)} = ((C-D) / (B-A) * \text{MS}) * 100$$

Dónde:

A = Peso del crisol

B = Peso del crisol con la muestra

C = Peso del crisol con el residuo desecado

D = Peso del crisol con las cenizas

MS = Materia Seca

3.3.7. DESARROLLO DE LOS MODELOS DE CALIBRACIÓN Y EVALUACIÓN DE LAS ECUACIONES.

Para el desarrollo de las ecuaciones de predicción NIRS se han utilizado las 100 muestras de los subproductos vegetales especificados en el apartado 1.. El primer paso en el desarrollo de los modelos de calibración fue obtener las ecuaciones de predicción mediante el software OPUS, versión 6 mediante el algoritmo PLS (mínimos cuadrados parciales), para obtener la relación entre el dato espectral y el

correspondiente valor químico. En estas primeras ecuaciones obtenidas no se realizó ningún tratamiento previo a los datos espectrales, sin embargo, se realizó la eliminación de muestras anómalas espectrales (OUTLIERS) mediante el cálculo de la distancia H (distancia de Mahalanobis) de cada muestra al centro de la población. Así se consideraron muestras atípicas aquellas que presentaron un valor de $H > 2$ (especificado en el programa OPUS).

Los espectros anómalos fueron eliminados repitiendo el proceso de calibración y eliminación de anómalos espectrales hasta lograr que todas las muestras presentaran un valor de H inferior al valor máximo recomendado. No eliminando más del 10% del conjunto de las muestras de calibración.

Posteriormente se procedió al desarrollo de los modelos de calibración obteniendo los parámetros estadísticos, los cuales, determinan la robustez del modelo de calibración/predicción. Esos parámetros estadísticos están definidos en el apartado 1.3.10 de la introducción y fueron los siguientes: R^2 calibración; RMSEE; RPD para la calibración y R^2 validación; RMSECV, RPD y el bias para la validación interna, indicando el número óptimo de factores a utilizar (Rank/rango), eligiendo aquel que presentó un RMSEE y RMSECV menor.

Una vez elegida la ecuación de calibración que presenta una mayor precisión en la estimación de cada uno de los componentes en muestras de subproductos vegetales, se pasó a realizar la optimización de la misma, utilizando los siguientes tratamientos estadísticos: Primera derivada, primera derivada con sustracción de línea, primera derivada y normalización vectorial con 17 puntos de suavizado, segunda derivada, corrección de dispersión multiplicativa (MSC), normalización vectorial (SNV) y normalización mínimos-máximos.

Una vez realizada la optimización se eligió aquella ecuación de calibración obtenida con el pretratamiento estadístico que presentó un mayor coeficiente de determinación (R^2_{cv}) y un menor RMSECV.

Posteriormente dicha ecuación estimara de forma precisa el porcentaje de grasa bruta, proteína bruta, fibra bruta, materia seca, materia orgánica y cenizas en muestras de subproductos vegetales.



4. RESULTADOS Y DISCUSIÓN

4.1. CARACTERÍSTICAS DE LAS MUESTRAS DE SUBPRODUCTOS VEGETALES UTILIZADOS PARA LA CALIBRACIÓN

En la **Tabla 4** se muestra la composición química del conjunto de las muestras de subproductos vegetales empleados en la calibración. En esta tabla se puede observar el número de muestras (N), la media, desviación estándar (SD), valores máximos y mínimos y el coeficiente de variación ($CV = SD/MEDIA$) para la grasa bruta, proteína bruta, fibra bruta, materia seca, materia orgánica, y cenizas.

Tabla 4. Contenido de los principales constituyentes en muestras de subproductos vegetales determinados por los métodos analíticos de referencia.

COMPONENTES	N	MEDIA	SD	MIN	MAX	CV
Grasa Bruta (%)	100	2,73	1,32	0,47	9,81	0,48
Proteína Bruta (%)	100	12,55	5,39	2,96	24,85	0,42
Fibra Bruta (%)	100	26,53	9,29	6,20	52,17	0,35
Materia Seca (%)	100	32,74	23,74	6,31	94,60	0,72
Materia Orgánica (%)	100	86,10	6,57	62,38	96,26	0,07
Cenizas (%)	100	13,89	6,57	3,73	37,61	0,47

N: Número de muestras; SD: Desviación estándar; MIN: Valores mínimos; MAX: Valores máximos; CV: Coeficiente de variación.

Para el componente de la grasa bruta, (Tabla 4) se obtuvo un valor medio de las muestras en la serie de calibración 2,73%. En el componente de la proteína bruta el valor medio es igual 12,55%. Los valores de la fibra bruta muestran un valor medio de 26,53% En el caso de la materia seca el valor medio obtenido es del 32,74%. Para la materia orgánica tenemos un valor medio 86,10%. Por último las cenizas se observa un valor de medio de 13,89%. Como ejemplo En el caso de la de la pulpa de naranja desecada según Bampidis et al. 2006 da unos valores medios muy similares en materia orgánica un 93,7% y

contenidos de proteína bruta en subproductos de alcachofa fue de 52,4% Sallam et al. 2008.

El amplio rango y la elevada desviación estándar del colectivo de calibración (**Tabla 4**) para cada uno de los componentes analizados y de forma particular para los constituyentes proteína bruta, fibra bruta, materia seca y cenizas confirma que dicho colectivo está constituido por subproductos vegetales muy diversos. Factor muy importante para obtener unas mejores ecuaciones de calibración y predecir posteriormente con mayor precisión y fiabilidad dichos componentes de forma rutinaria mediante el equipo de espectroscopía de infrarrojo cercano (NIRS).

El coeficiente de variación (CV), indica la desviación que existe con respecto al valor medio del constituyente. En general debido al amplio rango de los constituyentes se observa en la Tabla 4 valores altos para dicho parámetro estadístico. Siendo el coeficiente de variación (CV) más elevado para la materia seca debido a que presenta un rango mayor con respecto al resto de los componentes estudiados, mientras que el valor más pequeño corresponde a la materia orgánica, ya que la medida presenta una menor diferencia con respecto al mínimo y máximo establecido en el rango.

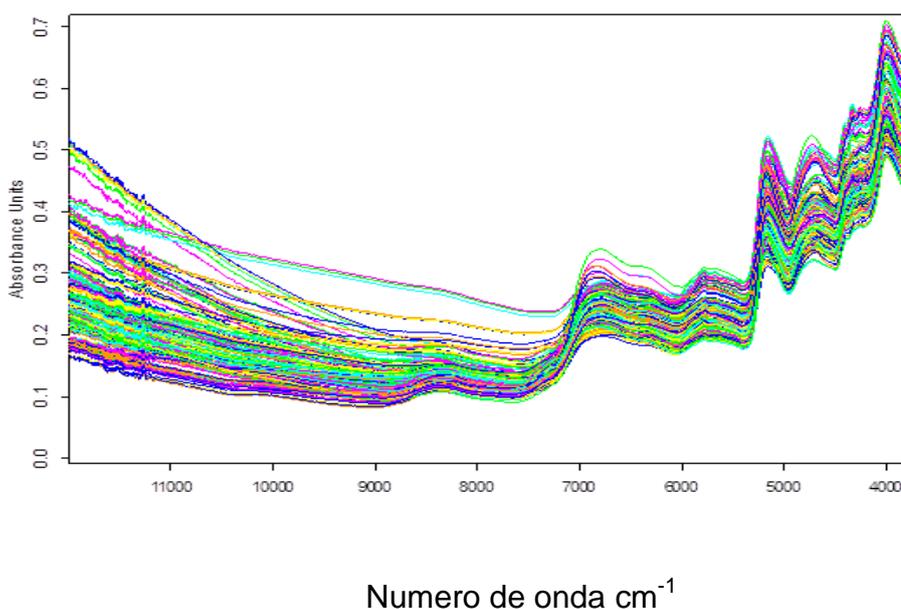
En la Figura 3 se observan los espectros sin pretratamiento previo de las 100 muestras de subproductos vegetales en la región del infrarrojo cercano en el rango de $12.500-4.000\text{ cm}^{-1}$ (800-2.500 nm). Los espectros de las muestras son similares en todo el rango espectral a lo largo del eje X. Si bien en el eje Y, se aprecian cambios en las intensidades espectrales entre las diferentes muestras de subproductos vegetales.

En los patrones espectrales NIR de las muestras fueron similares en toda la región de longitud de onda NIR $12,500-4000\text{ cm}^{-1}$ (800-2500 nm)

(Fig. 7) a lo largo del eje x. Si bien a lo largo del eje Y, los cambios de intensidades espectrales entre las diferentes muestras fueron claras.

En dicha figura se aprecian máximos de absorbancias para las longitudes de onda de 1.500 nm, 2.000 nm, 2.300 y 2.500 nm, estos máximos de absorbancia se deben en parte al hidrógeno que forma parte de los grupos funcionales $-CH$, $-OH$, y $-NH$ en los cuales el átomo de hidrógeno resulta ser el de menor masa favoreciendo las vibraciones de tensión o estiramiento (stretching) y las vibraciones de flexión (bending). El máximo de absorbancia a la longitud de onda de 2.300 nm, debido al grupo funcional $-CH$ puede ser apropiado para muestras de subproductos vegetales con bajo contenido en grasa (Murray et al., 1983). Autores como Osborne et al., 1993 asociaron las longitudes de onda 2352 nm con la celulosa, componente que forma parte de la fibra. También Clark y Lamb (1991) ponen de manifiesto que valores de absorbancia para las longitudes de onda comprendidas entre 2200-2300 nm están relacionadas con el grupo funcional $-CH$ de la fibra y particularmente con las vibraciones de tensión (stretching) de los grupos funcionales de las proteínas.

Figura 7. Conjunto de espectros de subproductos vegetales obtenidos en el rango $12.500-4.000\text{ cm}^{-1}$



4.2 DESARROLLO Y EVALUACIÓN DE LAS ECUACIONES DE CALIBRACIÓN SIN PRETRATAMIENTO ESPECTRAL.

A partir de los datos de referencia obtenidos por los métodos analíticos y los datos espectrales recogidos se pretende desarrollar ecuaciones de calibración para la determinación de grasa bruta (GB), proteína bruta (PB), fibra bruta (FB), materia seca (MS), materia orgánica (MO) y cenizas de manera in situ en muestras de subproductos vegetales. El método estadístico empleado para este objetivo fue el algoritmo de mínimos cuadrados PLS (mínimos cuadrados).

La **Tabla 5** muestra los estadísticos de calibración de las ecuaciones desarrolladas para cada parámetro analítico. Los mejores modelos para predecir la composición química de muestras de subproductos vegetales fueron aquellos que presentaron un menor valor para RMSEE y un valor mayor para R^2c (Windham et al., 1989). El número óptimo de factores fijado para calcular el modelo de calibración por PLS fue de 10 para todos los componentes excepto para la fibra bruta que fue de 9. Ese número óptimo de factores es aquel que produce la ecuación con el menor RMSEE.

Los valores del coeficiente de correlación (R^2c) (Tabla 5) de los parámetros analíticos evaluados para el colectivo de calibración fueron excelentes para la proteína bruta (R^2c) = 0.93), materia seca (R^2c) = 0.93) y aceptables para la grasa bruta (R^2c) = 0.70), fibra bruta (R^2c) = 0.74), materia orgánica (R^2c) = 0.89) y cenizas (R^2c) = 0.89). Este criterio considerado fue descrito por Shenk y Westerhaus, 1996. Es importante destacar que la amplia dispersión de los constituyentes (como puede verse en la desviación estándar de la Tabla 4) no ha proporcionado coeficientes de correlación inferiores a 0.6.

Los componentes proteína bruta, materia seca, materia orgánica y cenizas presentan unos RPDs (**Tabla 5**) superiores a un valor de 3, lo cual indican una buena predicción (Williams y Sobering, 1996). Sin embargo los RPDs de grasa bruta (1.85) y fibra bruta (1.98) se sitúan por

debajo de 3, estos valores nos indican que los modelos de predicción desarrollados para estos dos últimos componentes se podrían usar solo para indicarnos una estimación aproximada de su contenido en muestras de subproductos vegetales. También se observa en la **Tabla 5** que componentes con un valor menor de (R^2c) presentan un menor valor de RPD. Wang et al., 2014, encontraron valores similares para R^2c (0.94) y RPD (4.09) para la proteína bruta en muestras de soja.

Otro estimador para la capacidad predictiva de la ecuación es el error de calibración del modelo desarrollado (RMSEE) que 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) (Williams, 2001). En la **Tabla 5** los valores establecidos para el RMSEE en los componentes fibra bruta y materia seca presenta valores superiores a 4, lo cual es lógico, porque estos constituyentes presentan desviaciones estándar mayores que para el resto de los parámetros estudiados (**Tabla 4**). Para la proteína bruta, materia orgánica y cenizas son superiores a 1. Solo el componente grasa presenta un valor inferior a 1 (RMSEE=0.485), constituyente que presenta una menor variabilidad, con una desviación estándar de 1.32 (**Tabla 4**).

Xiccato et al., 1999, obtuvieron para la proteína bruta en alimentos destinados para la alimentación en conejos un valor elevado del RMSEE, siendo de 7.53. Este valor también estuvo asociado a una desviación estándar elevada para la proteína bruta (SD=21). Otros autores como Ferreira et al., 2013, para la grasa en muestras de soja obtuvieron en el modelo de calibración un valor ligeramente superior, 1.13 para el RMSEE.

Tabla 5. Valores de los parámetros estadísticos de las ecuaciones de calibración para los componentes analizados en los diferentes subproductos vegetales.

Componentes	Rango	Factores	R ²	RMSEE	RPD
GB (%)	0,93-5,96	10	0,709	0,48	1,85
PB (%)	2,96-24,85	10	0,936	1,35	3,97
FB (%)	6,91-52,17	9	0,744	4,32	1,98
MS (%)	4,03-94,60	10	0,931	4,81	3,81
MO (%)	62,38-96,26	10	0,892	1,89	3,05
CE (%)	3,73-37,61	10	0,891	1,95	3,04

GB: Grasa Bruta; PB: Proteína Bruta; FB: Fibra Bruta; MS: Materia Seca; MO: Materia Orgánica; CE: Cenizas; R²c: coeficiente de correlación

Sin embargo, el RMSEE, al ser únicamente un estimador del error del modelo quimiométrico empleado, dicho estadístico suele sobreestimar la capacidad real de predicción de una ecuación (Mark y Workman, 1991; Shenk Westerhaus, 1995b), de ahí la necesidad de realizar una validación interna llamada también validación cruzada y obtener un estimador más real, el RMSECV. De manera que cada muestra que forma parte del conjunto de calibración entrará en un pase de validación de una ecuación calculada a partir de las muestras restantes de calibración. El número de factores óptimos encontrado (mejor ecuación) será el de menor RMSECV (error típico de validación cruzada) medio entre todas las simulaciones.

En la **Tabla 6** se describen los estadísticos para evaluar la estimación de las ecuaciones de calibración realizando una validación interna llamada también validación cruzada. Como puede apreciarse los errores estándar de validación cruzada (RMSECV) fueron ligeramente inferiores al error estándar de calibración (RMSEE, **Tabla 5**), pero en ninguno de los constituyentes el valor de RMSECV fue inferior a el RMSEE multiplicado por 1.33 (Casal 2013). Esto indicaría una buena estimación de los componentes con las ecuaciones de calibración desarrolladas.

Los parámetros estadísticos mejores en la validación cruzada fueron para el componente de la proteína bruta con un alto valor de R²cv

(0.93) y un bajo RMSECV (1.43%) con 10 factores. Ferreira et al., 2013 obtuvieron valores para los mismos parámetros estadísticos con menor capacidad predictiva para estimar el contenido de proteína bruta en soja ($R^2_{cv} = 0.81$; $RMSEP = 1.61$).

Resultados similares se obtuvieron para el parámetro estadístico coeficiente de determinación (R^2) para la materia seca ($R^2_{cv} = 0.9136$), sin embargo el RMSECV fue superior, esto es debido a una mayor desviación estándar para el constituyente materia seca ($SD = 23.74$, **(Tabla 4)**, con respecto al componente de la proteína bruta ($SD = 5.39$, Tabla 4). Obteniendo además para estos últimos componentes un $RPD > 3$, lo cual, indica que la ecuación podría utilizarse para analizar cuantitativamente estos componentes con una elevada precisión (Williams y Sobering, 1996).

Para los componentes materia orgánica ($R^2_{cv} = 0.8242$; $RPD = 2.71$) y cenizas ($R^2_{cv} = 0.8644$; $RPD = 2.72$) obtenidos en el test de validación cruzada indican que van a proporcionar solo una buena precisión en los análisis de rutina para estos componentes

No obstante para el componente grasa bruta ($R^2_{cv} = 0.6076$; $RPD = 1.6$) y fibra bruta ($R^2_{cv} = 0.6938$; $RPD = 1.81$) las ecuaciones obtenidas para su predicción presentan una peor estimación que para el resto de los componentes. Según Belanche et al., 2014 establecen que valores de $R^2_{cv} < 0.67$ y $RPD < 2.08$ no son considerados suficientemente precisos para la estimación de los componentes.

Tabla 6. . Valores de los parámetros estadísticos de las ecuaciones de validación interna (validación cruzada) calculados para los componentes analizados en los diferentes subproductos vegetales.

Componentes	RANGO	FACTORES	R ²	RMSECV	RPD
GB (%)	0,93-5,96	10	0,607	0.54	1.6
PB (%)	2,96-24,85	10	0,925	1.43	3.66
FB (%)	6,91-52,17	9	0,693	4,61	1,81
MS (%)	4,03-94,6	10	0,913	5,24	3,4
MO (%)	62,38-96,26	10	0,864	2.06	2.71
CE (%)	3,73-37,61	10	0,864	2,12	2'72

GB: Grasa Bruta; PB: Proteína Bruta; FB: Fibra Bruta; MS: Materia Seca; MO: Materia Orgánica; CE: Cenizas; R²cv: coeficiente de determinación.

4.3. OPTIMIZACIÓN DE LAS ECUACIONES DE CALIBRACIÓN DESARROLLADAS Y EL ESTUDIO DE SU CAPACIDAD PREDICTIVA Y DE ESTIMACIÓN.

En la **Tabla 7** se observan los diferentes tratamientos estadísticos aplicados al conjunto de los espectros recogidos para optimizar la ecuación de calibración obtenida sin ningún tratamiento previo a las muestras, además de las regiones espectrales donde se obtiene el menor valor de RMSECV y el menor número de factores empleados para los valores de RMSECV obtenidos.

Los cinco tratamientos estadísticos que obtuvieron un menor RMSECV y números de factores empleados fueron primera derivada y segunda derivada utilizándose diferentes regiones del espectro y un quinto aplicando la primera derivada más sustracción de línea recta (**Tabla 7**).

Para el componente grasa (**Tabla 7**) en el tratamiento primera derivada las bandas de absorción se establecen en el rango de 7502,1-5446,3; 6102-5446,3; 4601,6-4246,7 cm⁻¹. Con el tratamiento primera derivada más sustracción de línea recta el rango se sitúa entre 11995-

4246,7 cm^{-1} . Con respecto a la segunda derivada el intervalo se establece entre 7502,1-4597,7 cm^{-1} .

Según Westad et al., 2008 y Williams, 2001 el componente grasa puede presentar pequeñas bandas de absorción a una longitud de onda de 1210 nm relacionados con el segundo overtone de la vibración stretching del enlace $-\text{CH}$, longitudes de onda 1726 y 1760 nm asociadas con el primer overtone de la vibración stretching del enlace $-\text{CH}$ y alrededor de 2308 y 2348 nm debido a la combinación de vibraciones stretching y bending del grupo funcional $-\text{CH}$.

El tratamiento espectral para optimizar el componente grasa que presenta mejores parámetros estadísticos en la ecuación de calibración es la segunda derivada, utilizando 10 factores con un $R^2c = 0.8196$; $\text{RMSEE} = 0,378$ y un $\text{RPD} = 2,35$ (**Tabla 7**). Sin embargo con respecto a los parámetros estadísticos de validación los más adecuados para estimar el constituyente grasa sería para el tratamiento de primera derivada ($R^2cv = 0,7498$; $\text{RMSECV} = 0,433$ y $\text{RPD} = 2$) utilizando 10 factores y seleccionando la región del espectro comprendida entre 7502,1-5446,3 cm^{-1} .

Según los valores reflejados en la **Tabla 7** el desarrollo de la ecuación de calibración para predecir el componente grasa en muestras de subproductos vegetales se debe aplicar el tratamiento espectral de primera derivada, utilizar como el número máximo de factores 10 en la ecuación de calibración y seleccionar el rango espectral comprendido entre 7502,1-5446,3 cm^{-1} .

Si se realiza una comparación con respecto a los estadísticos de calibración y validación obtenidos en las ecuaciones desarrolladas sin tratamiento previo a los datos espectrales ($R^2c = 0,794$; $\text{RMSEE} = 0,485$; $\text{RPD} = 1.85$; $R^2cv = 0,6076$; $\text{RMSECV} = 0,548$ y $\text{RPD} = 1,85$) (**Tabla 6**) se obtendría una mayor precisión en la estimación de dicho componente con la ecuación de calibración obtenida con el segundo tratamiento descrito en la **Tabla 7** ($R^2cv = 0,7498$; $\text{RMSECV} = 0.433$ y $\text{RPD} = 2$).

Tabla 7: Tratamientos espectrales, números de factores, regiones y parámetros estadísticos de calibración y de validación para el componente grasa.

OPTIMIZACION GRASA			CALIBRACIÓN			VALIDACIÓN			
Tratamiento	Rango	Regiones	R ²	RMSEE	RPD	R ²	RMSECV	RPD	Bias
2ª derivada	10	7502,1-5446,3; 4601,6-4246,7	0,819	0,37	2,35	0,716	0,46	1,88	-0,00106
1ª derivada	10	7502,1-5446,3	0,805	0,39	2,27	0,749	0,43	2	-0,000344
1ª derivada +sustracción	8	11995,-4246,7	0,796	0,39	2,22	0,712	0,46	1,86	0,00642
2ª derivada	9	7502,1-4597,7	0,796	0,40	2,22	0,732	0,44	1,93	-0,00268
1ª derivada	9	6102-5446,3; 4601,6-4246,7	0,776	0,42	2,11	0,720	0,45	1,89	0,00456

R²c: coeficiente determinación de calibración; R²cv coeficiente determinación de validación cruzada RPD: SD/RMSEE/RMSECV; RMSECV: Error típico de validación cruzada RMSEE: media de la raíz cuadrada del error de calibración/estimación;

Para la optimización de la ecuación de calibración desarrollada sin pretratamiento previo al conjunto espectral para estimar el componente proteína bruta (**Tabla 8**) se utilizaron primera derivada, primera derivada con sustracción de línea recta y normalización vectorial con 17 puntos de suavizado y segunda derivada.

Observando los diferentes tratamientos aplicados (**Tabla 8**), el pretratamiento estadístico aplicado al conjunto espectral para estimar de una forma fiable el componente proteína bruta fue la primera derivada, utilizando un máximo de 9 factores y seleccionando la región espectral comprendida entre dos rangos 11995,7-6098 cm⁻¹ y 5450,1-4246,7 cm⁻¹

Ferreira et al., 2013 utilizaron para la estimación del contenido de proteína bruta en soja la región del espectro comprendida entre 7500-4900 cm⁻¹ Sin embargo, Wang et al, 2014 establecen un rango en la región del espectro comprendido entre 11424,8-4242,8 cm⁻¹.en semillas de habas, valores muy similares a los establecidos en este trabajo. También Norris y Mark, 2004 establecieron que los rangos de absorción

para los grupos funcionales de la molécula de proteína se sitúan entre 1136-1334 y 1639-1836 nm.

Los parámetros estadísticos establecidos para el primer tratamiento (**Tabla 8**) en la optimización de la proteína bruta fueron para la ecuación de calibración $R^2c = 0,9577$; $RMSEE = 1,09$ y $RPD = 4,86$. No obstante, Ferreira et al., 2013 en muestras de soja encontraron un valor inferior para el coeficiente de determinación ($R^2c = 0,81$).

Sin embargo, para el estadístico $RMSEE$ encontrado por estos autores fue inferior ($RMSEE = 0,58$) al obtenido en este trabajo ($RMSEE = 1,09$). Esta diferencia es debida a la mayor desviación estándar que presentan las muestras del conjunto de calibración de este trabajo ($SD = 5,9$) con respecto a la desviación estándar encontrada por, Ferreira et al., 2013, estableciendo un valor de 2,34.

Con respecto a los parámetros estadísticos de validación ($R^2cv = 0,939$; $RMSECV = 1,28$ y $RPD = 4,05$ y con un bias de 0,0134) (Tabla 8) indican que la ecuación de calibración desarrollada para el componente proteína bruta va a estimar de forma precisa dicho componente en muestras de subproductos vegetales. Valores con menor capacidad estimativa fueron calculados por Ferreira et al, 2013, con un valor para $RMSECV$ de 1,61, un $RPD = 1,45$ y un bias $= -0,020$. Sin embargo Szgedi et al., 2012 en muestras de Brassica oleracea, aplicando el tratamiento de primera derivada con MSC (Corrección de dispersión multiplicativa) obtuvieron ecuaciones de calibración con parámetros estadísticos de mayor capacidad de estimación ($R^2c = 0,98$; $RMSECV = 0,76$).

Tabla 8: Tratamientos espectrales, números de factores, regiones y parámetros estadísticos de calibración y de validación para el componente de la proteína bruta

OPTIMIZACION PROTEINA BRUTA			CALIBRACIÓN			VALIDACIÓN			
Tratamiento	Rango	Regiones	R ²	RMSEE	RPD	R ²	RMSECV	RPD	Bias
1ª derivada	9	11995,7-6098,1; 5450,1-4246,7	0,957	1,09	4,86	0,939	1,28	4,05	0,0134
1ª derivada	9	11995,7-4246,7	0,955	1,12	4,74	0,936	1,31	3,96	-0,0122
1ª derivada + sustracción línea recta	9	11995,7-4246,7	0,953	1,15	4,61	0,929	1,38	3,77	-0,024
1ª derivada +normalización vectorial (SNV) puntos suavizado:17	9	11995,7-4246,7	0,951	1,17	4,54	0,927	1,4	3,71	0,0063
2ª derivada	9	9048,8-6098,1; 5450-4246,7	0,947	1,22	4,35	0,936	1,31	3,97	-0,0141

R²c: coeficiente determinación de calibración; R²cv coeficiente determinación de validación cruzada RPD: SD/RMSEE/RMSECV; RMSECV: Error típico de validación cruzada RMSEE: media de la raíz cuadrada del error de calibración/estimación; SNV: Standard Normal Variate;

Con respecto a la optimización de la ecuación de calibración desarrollada sin pretratamiento previo al conjunto espectral para estimar el componente fibra bruta (**Tabla 9**) se utilizaron los siguientes tratamientos estadísticos. corrección de dispersión multiplicativa (MSC); normalización vectorial (SNV); sustracción de línea recta; primera derivada y normalización vectorial; normalización mínimos-máximos.

Según los datos reflejados en la **Tabla 9**, de todos los pretratamientos estadísticos aplicados al conjunto espectral para estimar de una forma fiable el componente fibra bruta fue normalización vectorial (SNV), utilizando un máximo de 10 factores y seleccionando la región espectral comprendida entre dos rangos 11995,7-6098 cm⁻¹ y 4601,6-4246,7 cm⁻¹.

Belanche et al., 2014 exponen en su artículo que el rango comprendido entre 1050-1150 cm⁻¹. del espectro, puede ser debido a la vibración stretching de los enlaces –CO de la celulosa; -COC de la lignina y –CO de los taninos. Con respecto a las pectinas establecen que sus bandas de absorción se sitúan entre 1605-1630 cm⁻¹. Como se

puede observar son números de onda muy distintos a las obtenidas en este trabajo, incluso para los cinco tratamientos espectrales donde ponen de manifiesto el mismo rango de números de ondas en el espectro.

Los parámetros estadísticos establecidos para el tratamiento normalización vectorial (SNV) (**Tabla 9**) en la optimización de la fibra bruta fueron para la ecuación de calibración $R^2c = 0,780$; $RMSEE = 4$ y $RPD = 2,13$.

Con respecto a los parámetros estadísticos de validación ($R^2cv = 0,794$; $RMSECV = 4.47$ y $RPD = 1.85$ y con un bias de $-0,00095$) (Tabla 9) indican que la ecuación de calibración desarrollada para el componente fibra bruta estimaría con buena precisión dicho componente en muestras de subproductos vegetales.

Tabla 9: Tratamientos espectrales, números de factores, regiones y parámetros estadísticos de calibración y de validación para el componente de la fibra bruta.

OPTIMIZACION FIBRA BRUTA			CALIBRACIÓN			VALIDACIÓN			
Tratamiento	Rango	Regiones	R ²	RMSEE	RPD	R ²	RMSECV	RPD	Bias
Corrección de dispersión multiplicativa (MSC)	10	11995,7-6098,1; 5450,1-4246,7	0,784	3,96	2,16	0,704	4,51	1,84	0,025
Normalización vectorial (SNV)	10	11995,7-6098; 4601,6-4246,7	0,780	4,00	2,13	0,794	4,47	1,85	-0,00095
Sustracción de línea recta	10	11995,7-6098,1; 5450,1-4246,7	0,780	4,00	2,13	0,703	4,52	1,84	-0,0441
1ª derivada +SNV	7	11995,7-7498,3; 6102-4246,7	0,773	4,03	2,1	0,675	4,73	1,75	-0,0357
Normalización min-máx.	10	11995,7-6098; 4601,6-4246,7	0,764	4,14	2,06	0,697	4,57	1,82	-0,0108

R²c: coeficiente determinación de calibración; R²cv coeficiente determinación de validación cruzada RPD: SD/RMSEE/RMSECV; RMSECV: Error típico de validación cruzada RMSEE: media de la raíz cuadrada del error de calibración/estimación; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction

Con respecto a la optimización de la ecuación de calibración desarrollada sin pretratamiento previo al conjunto espectral para estimar

el constituyente de la materia seca (**Tabla 10**) se utilizaron los siguientes tratamientos estadísticos: primera derivada; corrección de dispersión multiplicativa (MSC); normalización vectorial (SNV) y segunda derivada.

Según los datos reflejados en la **Tabla 10**, de todos los pretratamientos estadísticos aplicados al conjunto espectral para estimar de una forma fiable el constituyente de la materia seca fue el tratamiento corrección de dispersión multiplicativa (MSC) utilizando un máximo de 10 factores y seleccionando la región espectral comprendida entre dos rangos 11995,7-4597,7 cm^{-1} y 8296,7-4597,7 cm^{-1} .

Ferreira et al., 2013 utilizaron para la estimación del contenido de materia seca en soja la región del espectro comprendida entre 10000-3500 cm^{-1} , rango espectral muy similar que el utilizado en este trabajo. Para la determinación de la materia seca es muy importante tener en cuenta las bandas típicas de absorción del agua en las longitudes de onda de 1450 y 1930 nm (Shenk et al., 1992). De ahí la importancia de realizar una desecación previa a las muestras empleadas en la calibración, con el objetivo que estas bandas de absorción del agua interfieran en lo menor posible en el análisis del resto de los componentes.

Los parámetros estadísticos establecidos para el tratamiento corrección de dispersión multiplicativa (MSC) (**Tabla 10**) en la optimización de la materia seca fueron para la ecuación de calibración $R^2c = 0,947$; $RMSEE = 4,16$ y $RPD = 4,36$.

Con respecto a los parámetros estadísticos de validación ($R^2cv = 0,932$; $RMSECV = 4,60$ y $RPD = 3,83$ y con un bias de $-0,075$) (**Tabla 10**) indican que la ecuación de calibración desarrollada para el componente materia seca estimaría con buena precisión dicho componente en muestras de subproductos vegetales.

Valores con menor capacidad estimativa para la materia seca fueron calculados en muestras de soja por Ferreira et al, 2013, con un

valor para R²_{cv} de 0,80, un RMSECV de 2,30, un RPD=1,38 y un bias =-0,049. Valores similares fueron encontrados por Xiccato et al., 2003 en alimentos destinados para la alimentación en conejos con un valor para R²_{cv} de 0,79, un RMSECV de 5,10.

Tabla 10: Tratamientos espectrales, números de factores, regiones y parámetros estadísticos de calibración y de validación para el componente de la materia seca.

OPTIMIZACION MATERIA SECA			CALIBRACIÓN			VALIDACIÓN			
Tratamiento	Rango	Regiones	R ²	RMSEE	RPD	R ²	RMSECV	RPD	Bias
1ª derivada	10	11995,7-3446,3	0,962	3,50	5,18	0,913	5,18	3,	-0,00642
Corrección de dispersión multiplicativa (MSC)	10	11995,7-4597,7	0,947	4,16	4,36	0,931	4,6	3,83	-0,075
SNV	10	11995,7-4597,7	0,942	4,36	4,16	0,925	4,80	3,67	-0,047
Corrección de dispersión multiplicativa (MSC)	10	8296,7-4597,7	0,939	4,45	4,07	0,930	4,65	3,79	-0,0605
2ª derivada	9	6102-4597,7	0,889	6,02	3	0,859	6,62	2,66	-0,0494

R²_c: coeficiente determinación de calibración; R²_{cv} coeficiente determinación de validación cruzada RPD: SD/RMSEE/RMSECV; RMSECV: Error típico de validación cruzada RMSEE: media de la raíz cuadrada del error de calibración/estimación; MSC: Multiplicative Scatter Correction

Con respecto a la optimización de la ecuación de calibración desarrollada sin pretratamiento previo al conjunto espectral para estimar el constituyente de la materia orgánica (**Tabla 11**) se utilizaron los siguientes tratamientos estadísticos: primera derivada con sustracción de línea recta y con normalización vectorial (SNV), segunda derivada y corrección de dispersión multiplicativa (MSC).

Según los datos reflejados en la **Tabla 11**, de todos los pretratamientos estadísticos aplicados al conjunto espectral para estimar de una forma fiable el constituyente de la materia orgánica fue la segunda derivada utilizando un máximo de 10 factores y seleccionando la región espectral en el rango comprendida entre 7502,1-4597cm⁻¹

La materia orgánica calculada como la diferencia de 100 menos el porcentaje de las cenizas pone de manifiesto que algunos autores indiquen que las ecuaciones de calibración para estimar dicho constituyente no presenten una buena estimación porque los minerales en forma pura, iónica o sales no absorben en la región del infrarrojo cercano (Shenk y Westerhaus, 1994). Sin embargo la ecuación de calibración podría ser más fiable en su estimación si los minerales están en las muestras en forma de complejos orgánicos o en forma de quelatos (Shenk y Westerhaus, 1994).

Los parámetros estadísticos establecidos para el tratamiento la segunda derivada (**Tabla 11**) en la optimización de la materia orgánica fueron para la ecuación de calibración $R^2c = 0,930$; $RMSEE = 1.52$ y $RPD = 3.78$.

Con respecto a los parámetros estadísticos de validación ($R^2cv = 0,930$; $RMSECV = 1,74$ y $RPD = 3,22$ y con un bias de $-0,011$) (**Tabla 11**) indican que la ecuación de calibración desarrollada para el componente materia orgánica estimaría con buena precisión dicho componente en muestras de subproductos vegetales. Valores con menor capacidad estimativa para la materia orgánica fueron encontrados por Xiccato et al., 2003 en alimentos destinados para la alimentación en conejos con un valor para R^2cv de 0,35 y un $RMSECV$ de 8,6.

Tabla 11: Tratamientos espectrales, números de factores, regiones y parámetros estadísticos de calibración y de validación para el componente de la materia orgánica.

OPTIMIZACION MATERIA ORGANICA			CALIBRACIÓN			VALIDACIÓN			
Tratamiento	Rango	Regiones	R ²	RMSEE	RPD	R ²	RMSECV	RPD	Bias
1ª derivada + sustracción de línea recta	10	11995,7-7498,3; 6102-4246,7	0,9403	1,41	4,09	0,8972	1,8	3,12	0,00182
1ª derivada + SNV	9	11995,7-7498,3; 5450,1-4246,7	0,9322	1,5	3,84	0,8855	1,90	2,96	-0,043
2ª derivada	10	7502,1-4597	0,9302	1,52	3,78	0,9307	1,74	3,22	-0,0115
1ª derivada	9	11995,7-6098,1; 5450,1-4246,7	0,9279	1,54	3,72	0,8893	1,86	3,01	-0,0312
Corrección de dispersión multiplicativa (MSC)	9	11995,7-7498,3; 5450,1-4246,7	0,9135	1,69	3,4	0,8866	1,89	2,97	-0,0112

R²c: coeficiente determinación de calibración; R²cv coeficiente determinación de validación cruzada RPD: SD/RMSEE/RMSECV; RMSECV: Error típico de validación cruzada RMSEE: media de la raíz cuadrada del error de calibración/estimación; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction

Con respecto a la optimización de la ecuación de calibración desarrollada sin pretratamiento previo al conjunto espectral para estimar el constituyente de las cenizas (**Tabla 12**) se utilizaron los siguientes tratamientos estadísticos: primera derivada, primera derivada con sustracción de línea recta, segunda derivada, corrección de dispersión multiplicativa (MSC), sustracción de línea recta.

Según los datos reflejados en la **Tabla 12**, de todos los pretratamientos estadísticos aplicados al conjunto espectral para estimar de una forma fiable el constituyente de las cenizas fue la segunda derivada utilizando un máximo de 10 factores y seleccionando la región espectral en el rango comprendida entre 7502,1-4597,7cm⁻¹.

Ferreira et al., 2013 utilizaron para la estimación del contenido de cenizas en soja la región del espectro comprendida entre 10000-3500 cm⁻¹, rango espectral muy similar que el utilizado en este trabajo.

Los parámetros estadísticos establecidos para el tratamiento la segunda derivada (**Tabla 12**) en la optimización de las cenizas fueron para la ecuación de calibración R²c =0,928; RMSEE=1.58 y RPD=3.75.

Con respecto a los parámetros estadísticos de validación R^2_{cv} =0,902; RMSECV=1,80 y RPD=3,20 y con un bias de 0,036) (**Tabla 11**) indican que la ecuación de calibración desarrollada para el componente las cenizas estimaría con buena precisión dicho componente en muestras de subproductos vegetales. Podemos observar que estos datos son muy similares a los encontrados para la ecuación de calibración desarrollada para estimar el porcentaje de la materia orgánica. Esto puede ser debido a la correlación entre ambos componentes.

Valores con menor capacidad estimativa para las cenizas fueron calculados en muestras de soja por Ferreira et al, 2013, con un valor para R^2_{cv} de 0,63, un RMSECV de 0,40, un RPD=0,95 y un bias =-0,080. Otros valores con peor capacidad predictiva fueron encontrados por Boever et al., 1995 en alimentos destinados para la alimentación en rumiantes con un valor para R^2_{cv} de 0,49.

Tabla 12: Tratamientos espectrales, números de factores, regiones y parámetros estadísticos de calibración y de validación para el componente de las cenizas

OPTIMIZACION CENIZAS			CALIBRACIÓN			VALIDACIÓN			
Tratamiento	Rango	Regiones	R^2	RMSEE	RPD	R^2	RMSECV	RPD	Bias
1ª derivada + sustracción de línea recta	9	11995,7-4246,7	0,930	1,55	3,80	0,895	1,87	3,09	0,00332
2ª derivada	10	7502,1-4597,7	0,928	1,58	3,75	0,902	1,80	3,20	0,0363
1ª derivada	9	11995,7-4246,7	0,927	1,59	3,72	0,890	1,91	3,02	0,020
Sustracción de línea recta	10	7502,1-6098,1; 5450,1-4246,7	0,927	1,74	3,39	0,884	1,96	2,94	0,00394
Corrección de dispersión multiplicativa (MSC)	9	11995,7-7498,3; 5450,1-4246,7	0,902	1,85	3,20	0,886	1,94	2,97	-0,0178

R^2_c : coeficiente determinación de calibración; R^2_{cv} coeficiente determinación de validación cruzada RPD: SD/RMSEE/RMSECV; RMSECV: Error típico de validación cruzada RMSEE: media de la raíz cuadrada del error de calibración/estimación; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction

5. CONCLUSIONES.

Según los valores de los parámetros estadísticos obtenidos en la validación interna o validación cruzada en la optimización de las ecuaciones de calibración desarrolladas para los componentes grasa bruta, proteína bruta, fibra bruta, materia seca, materia orgánica y cenizas, se puede concluir que:

- Para el componente grasa bruta la ecuación de calibración que estimaría con buena precisión dicho componente en muestras de subproductos vegetales sería la obtenida aplicando el pretratamiento de primera derivada al conjunto espectral, utilizando en la ecuación 10 factores máximos y seleccionando un rango espectral comprendido entre 7502,1-5446,3 cm^{-1} en el espectro.
- Para el componente proteína bruta la ecuación de calibración que estimaría con buena precisión dicho componente en muestras de subproductos vegetales sería la obtenida aplicando el pretratamiento de primera derivada al conjunto espectral, utilizando en la ecuación 9 factores máximos y seleccionando dos rangos espectrales comprendidos entre 11995,7,1-6098,1 cm^{-1} y 5450,1-4246,7 cm^{-1} en el espectro.
- Para el componente fibra bruta la ecuación de calibración que estimaría con buena precisión dicho componente en muestras de subproductos vegetales sería la obtenida aplicando el pretratamiento de normalización vectorial (SNV) al conjunto espectral, utilizando en la ecuación 10 factores máximos y seleccionando dos rangos espectrales comprendidos entre 11995,7-6098 cm^{-1} y 4601,6-4246,7 cm^{-1} en el espectro.
- Para el componente materia seca la ecuación de calibración que estimaría con buena precisión dicho componente en muestras de subproductos vegetales sería la obtenida aplicando el

pretratamiento de corrección de dispersión multiplicativa (MSC) al conjunto espectral, utilizando en la ecuación 10 factores máximos y seleccionando un rango espectral comprendido entre 11995,7-4597,7 cm^{-1} en el espectro.

- Para el componente materia orgánica la ecuación de calibración que estimaría con buena precisión dicho componente en muestras de subproductos vegetales sería la obtenida aplicando el pretratamiento de segunda derivada al conjunto espectral, utilizando en la ecuación 10 factores máximos y seleccionando un rango espectral comprendido entre 7502,1-4597 cm^{-1} en el espectro.

- Para el componente cenizas la ecuación de calibración que estimaría con buena precisión dicho componente en muestras de subproductos vegetales sería la obtenida aplicando el pretratamiento de segunda derivada al conjunto espectral, utilizando en la ecuación 10 factores máximos y seleccionando un rango espectral comprendido entre 7502,1-4597,7 cm^{-1} en el espectro.

6. BIBLIOGRAFIA.

AOAC (1995). Official methods of analysis of AOAC International. 16th ed. Washington (D.C): AOAC International.

ARRUABARRENA, I. A. (2012). Evaluación de calidad de la patata mediante tecnología NIRS y la influencia del sistema productivo convencional y ecológico. UPNA, Pamplona.

ASTM, (1994). Standard practices for infrared, multivariate, quantitative analysis.

BARNES, R.J., DHANOA M.S. Y LISTER, S.J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43 pp. 772-777.

BEEBE, K.R., PELL, R.J. Y SEASHOLTZ, M.B. (1998). Chemometrics. A practical guide, John Wiley & sons, New York.

BELANCHE, A., WEISBJERG, M. R., ALLISON, G. G., NEWBOLD, C. J., & MOORBY, J. M. (2014). Measurement of rumen dry matter and neutral detergent fiber degradability of feeds by Fourier-transform infrared spectroscopy. Journal of dairy science, 97(4), 2361-2375.

BEN-GERA, I. Y NORRIS K.H. 1968. Direct spectrophotometric determination of fat and moisture in meta products. Journal of Food Sciences 7 p. 240.

BERTRAND, D. (1993). Data pre-treatment and original analysis in spectroscopy. Advanced Comet Chemometrics School Libramont Belgium, April 26-28, 1993.

BERTRAND, D. Y DUFOUR, E. (2000). La spectroscopie infrarouge et ses applications analytiques. Editions TEC & DOC, Paris, France.

BURNS, D.A. Y CIURCZAK, E.W. (1992). Handbook of near infrared analysis. Burns, D. A. and Ciurczak, E.W. (eds.) Marcel Dekker, New York, NY, USA.

DE BOEVER, J. L., COTTYN, B. G., VANACKER, J. M., & BOUCQUÉ, C. V. (1995). The use of NIRS to predict the chemical composition and the energy value of compound feeds for cattle. *Animal Feed Science and Technology*, 51(3), 243-253.

CÁCERES, P. 1999. FUNDAMENTOS NIR. Curso de Bran +Luebbe. S/P CEN, H. Y HE, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science and Technology*, 18, 72-83.

CLARCK, D.H AND LAMB, R.C 1991. Near infrared spectroscopy: a survey of wavelength selection to determine dry matter digestibility. *Journal Dairy science* 74, 2200-2205

D.H CLARK, R.C LAMB 1991. Near infrared reflectance Spectroscopy: a study selection wavelength to determine dry matter digestibility *J. Dairy Sci*, 74 (1991), pp. 2200-2205

DAVIES, A.M. Y GRANT, A. (1987). Review: Near infra-red analysis of food. *International Journal of Food Science and Technology*, 22, 191-207.

DEAVILLE, E.R. Y FLINN, P.C. (2000). Forage Evaluation in Ruminant. pp. 301-320.

DHANOVA, M.S., LISTER, S.J., SANDERSON, R. Y 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.

FERNÁNDEZ-CABANÁS, V Y 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.

FERREIRA, D. S., PALLONE, J. A. L., & POPPI, R. J. 2013. Fourier transform near-infrared spectroscopy (FT-NIRS) application to estimate Brazilian soybean [*Glycine max* (L.) Merrill] composition. *Food research international*, 51(1), 53-58.

GARRIDO (1997) Current and future applications of NIRS technology in the feed industry in Morand-Fehr P. (ed.) . *Feed manufacturing in Southern Europe: New challenges* pp, 87-92

GARRIDO, A., GÓMEZ, A., GUERRERO, J. E., & FERNÁNDEZ, V. (1996). NIRS: Una tecnología de apoyo para un servicio integral en Alimentación Animal. *Avances en Alimentación y Nutrición Animal*, 275-300.

HENRY, C.M. (1999). Near-IR gets the job done. *Analytical Chemistry* 71 17 p. 625A-628A.

HORLICK, G. (1972). Digital data handling of spectra utilizing Fourier transformations. *Anal. Chem.* 44 p. 943-947.

I. N. R. A., 1981. *Prévision de la valeur nutritive des aliments des ruminants*. Ed.: I.NR.A publications. Versailles (France). 590 pp.

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. pp. 175-186.

JARÉN, C.; ORTUÑO, J. C.; ARAZURI, S. y ARANA, J. I. (2001). Sugar determination in grapes using NIR technology. *International Journal of Infrared and millimeter waves*, 22 (10). Págs. 1521, 1522.

MAGRAMA, 2014. Anuario de estadística. en.

http://www.magrama.gob.es/estadistica/pags/anuario/2014Avance/AE_2014_Avance.pdf

MARK AND WORKMAN, 1991; H. Mark, J. Workman; *Statistics in spectroscopy*, Academic Press, San Diego (1991), pp. 263–302.

MARK, H. (2004). Qualitative Near-Infrared Analysis. In: *Nearinfrared technology in the agricultural and food industries*. Chapter 13 (edited by Norris, K. & Williams, P.) pp. 233–234. St Paul, MN, USA: American Association of Cereal Chemists, Inc.

MARTENS, H. NAES, T. (1989). *Multivariate calibration*. John Wiley and Sons Inc., Chichester, UK.

MARTENS, H. Y 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 KAUFMAN, L. (1988). *Chemometrics: a textbook*. Elsevier Science Pub., Amsterdam, the Netherlands.

MASSART, D.L.; VANDEGINSTE, B. G. M.; BUYDENS, L. M. C. (1988). *Chemometrics: A textbook*. Vol. 2. Ed. Elsevier, Amsterdam.

MCDONALD, P. 1981. *The biochemistry of silage*. U.K., J. Wiley. 226 p.

MILLMIER A; LORIMOR J; HURBURGH JR, C; FULHAGE C; HATTEY J; ZHANG H (2000). Near-infrared sensing of manure nutrients. *Transactions of the ASAE*, 43 (4), 903–908.

MURRAY, I., HALL, P.A., (1983). Animal feed evaluation by use of near infrared reflectance (NIR) spectrocomputer. Anal. Proc. 20, 75-79.

MURRAY, P.A HALL (1983) Animal feed evaluation by use of near infrared reflectance (NIR) spectrocomputer Anal. Proc., 20 (1983), pp. 75–79.

NAES T, ISAKSSON, T., FEARN, T. DAVIES, T. (2002). A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK.

NAES, T., ISSAKSSON, T., FEARN, T. Y DAVIES, T. (2002). A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK.

NAES, T., ISSAKSSON, T., FEARN, T. Y DAVIES, T. (2002). A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK.

NICOLAÏ, B.M., BEULLENS, K., BOBELYN, E., PEIRS, A., SAEYS, W., THERON, K.I Y LAMMERTYN, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology, 46, 99-118.

NORRIS, K. H. (1964). Simple spectroradiometer for 0.4 to 1.2- micron region. Trans. ASAE 7(3), pp. 240- 242.

NORRIS, K.H., WILLIAMS, P.C.(1984).Optimization of mathematical treatments of raw nearinfrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size. Cereal Chemistry 61 2 pp. 158-165.

OJEDA, F.; CÁCERES, O.; ESPERANCE, M. (1991). Conservación de Forrajes. Editorial Pueblo y Educación. 80p.

OSBORNE, B.G. Y FEARN, T. (1986). Near Infrared Spectroscopy in Food Analysis Longman Scientific and Technical, Halow, Essex, UK.

OSBORNE, B.G., FEARN, T. Y HINDLE, P. (1993). Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Longman Scientific and Technical London, UK.

OSBORNE, B.G., FEARN, T. Y HINDLE, P. (1993). Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Longman Scientific and Technical London, UK.

OSBORNE, B.G., FEARN, T., HINDLE, P.H., (1993). Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Longman Scientific and Technical, Essex CM 2JE, England, pp. 227.

P.C WILLIAMS, P.M STARKEY Influence of feed ingredients upon the prediction of protein in animal feed-mixes by near-infrared reflectance spectroscopy J. Sci. Food Agric., 31 (1980), pp. 1201–1213

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. Tesi Doctoral. Universitat de Barcelona.

SAVITZKY A. AND GOLAY, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36 pp. 1627-1639.

SZIGEDI, T., LÉNÁRT, J., DERNOVICS, M., TURZA, S., & FODOR, M. (2012). Protein content determination in Brassica oleracea species using FT- NIR technique and PLS regression. International Journal of Food Science & Technology, 47(2), 436-440.

SHENK, J.S AND WESTERHAUS, M.O. (1991). 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. WESTERHAUS, M.O. (1996). Calibration the ISI way. *Near Infrared Spectroscopy: The Future Waves*. Davies, A.M.C. and Williams, P.C. (eds.). NIR Publications, Chichester, UK. pp. 198-202.

SHENK, J.S. Y WESTERHAUS, M.O. (1995a). *Analysis and Food Products by Near Infrared Reflectance Spectroscopy*. Monograph. NIRSystems Inc., Silver Spring, MD, USA.

SHENK, J.S. Y WESTERHAUS, M.O. (1995). *Analysis and Food Products by Near Infrared Reflectance Spectroscopy*. Monograph. NIRSystems Inc., Silver Spring, MD, USA.

SHENK, J.S., WESTERHAUS, M.O., 1995b. *Routine Operation, Calibration, Development and Network System Management Manual*. NIRSystem, Silver Spring, MD.

SHENK, J.S., WORKMAN, J.J. AND WESTERHAUS, M.O. (1992). *Application of NIR Spectroscopy to Agricultural Products*. En *Handbook of near infrared analysis*.

SKOOG, D.A. y LEARY, J. J. (1996). *Análisis instrumental*. Ed. McGraw-Hill/Interamericana de España S. A., Madrid, España. Págs.: 296-339.

STONE, M. (1974). Cross validation choice and assessment of statistical prediction. *J. R. Statist. Soc. B*, 39:111-133.

W.R. WINDHAM, D.R. MERTENS, F.E. BARTON (1989), *II Protocol for NIRS calibration: Sample selection and equation development and validation* C.G. Marten, J.S. Shenk, F.E. Barton (Eds.), *Near infrared*

reflectance spectroscopy (NIRS): Analysis of forage quality (643rd ed.), USDA-ARS: Agricultural Handbook, Washington, DC pp. 96–103

WANG J., LIU H., REN G., (2014) Near-infrared spectroscopy (NIRS) evaluation and regional analysis of Chinese faba bean (*Vicia faba* L.) The crop Journal (2014) 28-37

WESTAD, F., SCHMIDT, A., & KERMIT, M. (2008). Incorporating chemical band-assignment in near infrared spectroscopy regression models. *Journal of Near Infrared Spectroscopy*, 16(3), 265-273.

WILLIAMS, P C Y 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 futures waves. Davies, A.M.C.and Williams, P. c. (Eds). NIR Publications. Chischester. UK 185-188.

WILLIAMS, P. 2001. Near-infrared technology: In the agricultural and food industries. In P. Williams, & K. Norris (Eds.), (2nd ed.). American Association of Cereal Chemists. pp. 296) St. Paul, Minn., USA: American Association of Cereal Chemists.

WILLIAMS, P., (2003). Near-infrared Technology–Getting the Best out of Light. PDK Grain. Nanaimo, Canada.

WILLIAMS, P.C. Y NORRIS, K., (1987). Near-infrared technology in the agricultural and food industries. Williams, P.C. and Norris, K. (eds.) American Association of Cereal Chemist Inc., St. Paul, MN, USA.

WILLIAMS, P.C. Y NORRIS, K., (1987). Qualitative applications of nearInfrared Reflectance Spectroscopy. American Association of Cereal Chemist Inc., St Paul, MN, USA. p. 241-246.

WOLD S., SJÖSTRÖM, M. Y ERIKSSON, L. (2001). PLS- regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58 (2001) 109–130.

XICCATO, G., TROCINO, A., DE BOEVER, JL, MAERTENS, L., CARABAÑO, R., PASCUAL, JJ, & FALCAO-E-CUNHA, L. (2003). Prediction of chemical composition, nutritive value and ingredient composition of European compound feeds for rabbits by near infrared reflectance spectroscopy (NIRS). *Animal Feed Science and Technology*, 104 (1), 153-168.

XICCATO G., TROCINO, A., CARAZZOLO A., MEURENS M., MAERTENS L., CARABAÑO R., 1999. Nutritive evaluation and ingredient prediction of compound feeds for rabbits by near-infrared reflectance spectroscopy (NIRS). *Animal Feed Science and Technology* 77 201-212.

CASAL MARTÍNEZ, C. (2013). Implantación de la tecnología NIRS en aplicaciones in-situ como herramienta de apoyo a las decisiones en el sector primario.

Garrido, A., Gómez, A., Guerrero, J. E., & Fernández, V. (1996). NIRS: Una tecnología de apoyo para un servicio integral en Alimentación Animal. *Avances en Alimentación y Nutrición Animal*, 275-300.

WINDHAM, W. R., MERTENS, D. R., & BARTON, F. E., II (1989). Protocol for NIRS calibration: Sample selection and equation development and validation. In C. G. Marten, J. S. Shenk, & F. E. Barton (Eds.), *Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality* (pp. 96–103). (643rd ed.). Washington,DC: USDA-ARS: Agricultural Handbook.

BAMPIDIS, V. A., & ROBINSON, P. H. (2006). Citrus by-products as ruminant feeds: A review. *Animal Feed Science and Technology*, 128(3), 175-217.

SALLAM, S. M. A., BUENO, I. C. S., GODOY, P. B., NOZELLA, E. F., VITTI, D. M. S. S., & ABDALLA, A. L. (2008). Nutritive value assessment of the artichoke (*Cynara scolymus*) by-product as an alternative feed resource for ruminants. *Tropical and Subtropical Agroecosystems*, 8(2), 181-189.

