INFORME TÉCNICO SEGUIMIENTO

PROGRAMA ASTURIAS

DATOS DE IDENTIFICACIÓN DEL PROYECTO DE I+D+i				
Referencia proyecto	IDI/2018/000095			
Acrónimo	NIRFORIND			
Título del Proyecto	Desarrollo de nuevas aplicaciones de la espectroscopia NIR orientadas a la industria agroalimentaria asturiana.			
Periodo de Justificación	Desde :	01/01/2018	Hasta:	31/12/2020

En Noreña , a 24 de marzo de 2021

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

1. INFORMACIÓN ECONÓMICA

Explique y justifique las desviaciones o los cambios producidos en las diferentes partidas del coste subvencionable aprobado, no sometidos a autorización expresa e indique su impacto en el desarrollo del proyecto.

No se ha producido ninguna desviación ni cambio.

2. INCIDENCIAS, MODIFICACIONES Y/O DESVIACIONES

¿Se ha realizado el proyecto conforme a la solicitud presentada?: **SI** ☐ **NO** ☒

En caso negativo, especifique las incidencias, modificaciones y/o desviaciones, explicando las causas y detallando cómo han afectado al normal desarrollo del proyecto.

En el año 2019, se ha producido un retraso en la obtención de muestras de diferentes especies de túnidos ya que las empresas productoras no han podido colaborar. Se analizaron varias medidas correctoras que fueron:

- Esperar a otra época del año con menor carga de trabajo para las empresas
- Se valora, si el primer punto no funciona, la recogida de muestras de producto final (latas). Ya hemos hecho en 2019 alguna prueba previa
- Se incrementó el trabajo en las otras dos matrices con aumento del número de muestras de café y kiwi

En el año 2020 no se ha producido ningún incidente. Se utilizaron latas y botes de atún para la clasificación de especies.

3. MEMORIA TÉCNICA

Explique **DETALLADAMENTE** las actividades realizadas, incluyendo la descripción de la metodología empleada, con referencia expresa a los hitos, tareas y plan de trabajo de la solicitud.

HITO 1.- ESPECIFICACIONES TÉCNICAS DE USUARIO.

TAREA 1.1.- ESPECIFICACIONES DE LAS MATRICES ALIMENTARIAS.

CAFÉ:

Para el caso del café, se estudiarán las siguientes propiedades:

Estimación de la composición de mezclas para Arábica/robusta:

Las dos variedades de café más comunes a nivel mundial son la Arábica y la Robusta. Existen diferencias significativas entre ambos cafés. Entre ellas cabe destacar es que el café Arábica se cultiva a gran altitud, mientras que el café Robusta se cultiva en zonas más bajas. El 70 % de la producción mundial es de café Arábica y el 30 de Robusta. En México, Centroamérica, zona pacífica de américa del sur, y países africanos como Etiopía y Kenia, se cultiva mayoritariamente café Arábica. El cultivo de la variedad de café Robusta es mayoritario en países de Asia, Oceanía y en zonas de áfrica occidental y central. En lugares como Brasil, África Central, África Oriental y Madagascar se cultivan extensamente las dos variedades. Los granos de café arábica son más alargados y más verdosos, mientras que los de arábica

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACIÓN PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACIÓN CIENTÍFICA APLICADA Y LA TECNOLOGÍA
iasturias INNOVACIÓN

son más redondos y amarillentos. El café Arábica es menos amargo que el café Robusta. Además, el café Arábica es más perfumado y tiene mejor sabor y olor. En pocas palabras se puede afirmar que el café Arábica está considerado de mejor calidad, pero su precio es mayor que el café Robusta.

A nivel genético la primera diferencia es que el café Robusta es un diploide (posee dos series de cromosomas) y la variedad Arábica es un tetraploide (posee cuatro pares de cromosomas), siendo la única especie con esta característica[1]. (Belitz et al., 2009)[2] exponen las diferencias para distintos compuestos para las variedades Arábica y Robusta de café verde. Para la cafeína, se puede ver que para la variedad Arábica tiene entre un 0.8 % y un 1.4 %, mientras que el café Robusta tiene una cantidad mayor que se suele encontrar en el intervalo 1.7-4 %. El contenido en metales[3] es otro aspecto diferenciador, siendo el fósforo, el manganeso y el cobre, los métales que muestran mayores diferencias en su contenido para las dos variedades de café. También, la fracción esterólica[4] del aceite de las mezclas puede ser usada para establecer el porcentaje de Arábica/Robusta. El Δ^5avenasterol es el químico descriptor más adecuado para predecir el contenido de Arábica/Robusta. En otro trabajo[5], se muestra la utilidad del uso del tocoferol para discernir entre las dos variedades de café tanto en café verde como tostado. Por otra parte, se pueden usar determinados marcadores basados en la composición de los ácidos grasos[6] para diferenciar entre los dos tipos de café. A diferencia de otros estudios que usan marcadores que se ven fuertemente afectados por el grado de tostado que influencia fuertemente los resultados, se usan los ácidos linoleicos, α-linoleicos que son más abundantes en el café Arábica, mientras que el ácido oleico presenta cantidades mayores en el café Robusta. (Bicho et al., 2013)[7] estudiaron la aplicabilidad de diferentes parámetros químicos para discernir entre café verde Arábica y Robusta. En general, el café Robusta presenta valores mayores de pH, ácidos solubles, cafeína, total de ácidos cafeilquínicos, total de ácidos dicafeilquínicos y ácido feruloilquínico. A través de la aplicación de análisis multivariado, donde se usan funciones de distancia/disimilaridad, se ha encontrado que la cafeína, trigonelina, ácido 3-O-cafeicoilquínico, ácido 4-O-cafeicoilquínico, ácido 3-O-feruloilquínico, ácido 5-O-feruloilquínico, ácido 3,4-dicafeicoilquínico, ácido 3,5-dicafeicoilquínico y ácido 4,5-icafeicoilquínico pueden ser considerados un cluster de discriminación química entre las variedades de café verde Arábica y Robusta.

Humedad Café:

La humedad en el café se estudiará para el café verde y el café tostado, tanto en grano como molido. La razón es porque la humedad en el café es un factor clave en todo el proceso de producción de café, ya que la presencia de humedad en cualquier lugar del proceso productivo genera efectos adversos como pueden ser:

- En café verde, un contenido de agua superior al 12.5 % provoca aspectos adversos como crecimiento microbiano, formación de micotoxinas o alteración de la cualidad sensorial. Además, influye en gran medida en el tostado del café. Esto se debe a que cuanta más humedad tenga el grano verde, más energía se necesitará para tostar el mismo. Además, otro factor

[1] Pinto-Maglio, C.A.F. Cytogenetics of coffee. *Brazilian Journal of Plant Physiology*. 18 (2006) 37-44.
[2] Belitz, H.D.; Grosch, W.; Schieberle, P. Food Chemistry. Springer. 2009.
[3] Martín, M.J.; Pablos, F.; Gonzalez, A.G. Characterization of arabica and robusta roasted coffee varieties and mixture resolution according to their metal content. *Food Chemistry*. 66 (1999) 365-370.
[4] Valdenebro, M.S.; León-Camacho, M.; Pablos, F.; Gonzalez, A.G.; Martín, M.J. Determination of the arabica/robusta composition of roasted coffee according to their sterolic content. *The Analyst*. 124 (1999) 999-1002.
[5] Alves, R.C.; Casal, S.; Alves, M.R.; Oliveira, M.B. Discrimination between arabica and robusta coffee species on the basis of their tocopherol profiles. *Food Chemistry*. 144 (2009) 295-299.
[6] Romano, R.; Santini, A.; Le Grottaglie, L.; Manzo, N.; Visconti, A.; Ritieni, A. Identification markers based on fatty acid composition to differentiate between roasted Arabica and Canephora (Robusta) coffee varieties in mixtures.
[7] Bicho, N.C.; Leitão, A.E.; Ramalho, J.C.; de Alvarenga, N.B.;Lidon, F.C. Identification of chemical clusters discriminators of Arabica and Robusta Green coffee. *International Journal of Food Properties*. 16 (2013) 895-904.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias Innovación

económico derivado de este aspecto, es que cuanta más humedad tenga, este café tendrá porcentualmente menos café, con la consiguiente pérdida de dinero. También, una alta humedad influye en el aroma de los tuestes ligeros. En café verde, la humedad suele estar entre un 8% y un 12.5%, pero se considera que una humedad en torno a 10-12% es un rango bueno para un café de calidad.

- Para el café tostado, tanto en grano como molido, una alta humedad conlleva peores cualidades sensoriales. Además, el café molido tiene aspecto apelmazado cuando presenta humedad. Estos dos aspectos conllevan una disminución de la aceptación de los clientes y descenso en las ventas. La humedad del café tostado cuando sale de la tostadora suele estar entre un 2 y un 3 %. Diferentes problemas asociados con el proceso de tostado, pueden conllevar desviaciones en este rango.

Parámetros relativos al tostado del café:

El resto de parámetros relativos al tostado del café dependen del tipo de tostado. Así un tueste ligero produce un café más claro, y según se aumentan los tiempos y temperaturas de tostado se obtienen tuestes más oscuros. Esta diversidad es consecuencia de los gustos de cada región. Así, por ejemplo, en Estados Unidos suelen tomar el café poco tostado y en Italia se toma mucho más tostado. En la misma situación que el color se encuentran la pérdida de peso y la densidad. Estos dos parámetros dependen altamente del tipo de tostado, también dependen de las condiciones de tueste. La pérdida de peso es originada por la pérdida de humedad y materia orgánica. Cuanto mayor sea el tostado mayor pérdida de peso. En función del tipo de tostado se puede perder entre un 12 y un 20 % del peso. El cambio en la densidad de los granos es función del aumento de volumen de los granos y la pérdida de peso. A mayor tostado, menor densidad de los granos de café. La densidad del café verde es alrededor de 1.25 gr/cm^3, la del café con tueste de ligero a medio suele ser de entre 0.5-0.75 gr/cm^3, y la del café con tueste de medio a oscuro de 0.4-0.6 gr/cm^3.

ATÚN:

La problemática en la identificación de atunes es resultado del diferente precio que tienen las distintas especies en el mercado.

Las cuatro especies que es habitual encontrar en una captura[8] son:

- Bonito (*Thunnus Alalunga*).
- Skipjack (*Katsuwomus Pelanis*).
- Yellowfin (*Thunnus Albacanes*).
- BigEye (*Thunnus Obesus*).

Actualmente, expertos realizan la clasificación en crudo por diferencias morfológicas entre peces. Cuando los atunes son recién capturados son relativamente fáciles de distinguir. Cuando los atunes llegan a una sala de procesado, la identificación es más complicada. Existen varios escenarios donde es altamente complicado identificar la especie:

- En las tallas pequeñas es muy difícil distinguir entre Yellowfin y BigEye.

[8] Capturas nominales mundiales del atún. Colecciones de estadísticas de pesca. Organización de las Naciones Unidas para la Alimentación y la Agricultura. Departamento de Pesca y Acuicultura. 2010.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

- Cuando el pescado presenta deformaciones que modifican las diferencias morfológicas en las que se basan los expertos para la identificación.
- Los atunes empiezan un proceso de descoloración a partir de la muerte. El proceso de descoloración es más acusado si el atún se congela.
- Cuando el atún es cocido y limpiado, su identificación solo es posible con ADN.

La composición bioquímica de las diferentes especies de atún queda patente en la bibliografía científica[9,10]. Así, entre dos de las especies (Yellowfin y Bigeye) que presentan mayores problemas en su separación en planta, se presentan diferencias en su bioquímica, como diferentes proporciones de ácido glutámico, metionina y en diferentes ácidos grasos. También el Skipjack y el Yellowfin presentan diferentes porcentajes de ácidos grasos saturados, ácidos grasos monoinsaturados y ácidos grasos polinsaturados para los diferentes puntos de su anatomía.

El objetivo ideal para las empresas del sector, es tener un método que logre distinguir especies en pescado cocido. Es especialmente interesante la distinción entre Yellowfin-Bigeye.

KIWI:

Para el caso del estudio del kiwi asturiano, se abarcará desde la etapa de crecimiento de este en el árbol hasta su maduración en cámara frigorífica. Se estudiarán kiwis de diferentes productores asturianos y se abarcará las etapas de crecimiento posteriores a la recolección, y su almacenamiento en cámara frigorífica.

En concreto se estudiará el kiwi verde asturiano, que es una baya ovoide con piel de color marrón/verdoso y recubierto de pelo. En el interior, la pulpa es de color verde y sus semillas se agrupan alrededor del corazón blanco central. Un kiwi comercial pesa entre 80-120 g[11].

En Asturias, los kiwis se suelen recolectar a principios de noviembre. Según el Reglamento (CE) Nº 1221/2008 de la Comisión de 5 de diciembre de 2008, "*en la fase de acondicionamiento en la región de producción y en la entrega siguiente efectuada por el envasador, así como en las fases de exportación y de importación, de por lo menos 6,2 ºBrix o un 15 % de contenido medio de materia seca*". Además, según este mismo reglamento "*en todas las demás fases de comercialización, de por lo menos 9,5 ºBrix*".

Para este proyecto estudio se recolectarán los kiwis antes de alcanzar los requisitos de recolección que marca el reglamento Nº 1221/2008. Para ello se recolectarán kiwis del árbol en fechas anteriores al mes de noviembre, en el mes de octubre. Además, para estudiar la maduración del kiwi posterior a la cosecha, una vez recolectados a principios de noviembre, se almacenarán y se analizarán en tandas temporales que abarquen su maduración postcosecha, hasta alcanzar estados de maduración avanzada.

Teniendo todo esto en cuenta, los grados Brix se moverán en un rango desde 3 o 4 ºBrix hasta unos 15 ºBrix. La acidez tiene un rango de 0.5 hasta 3 % en acidez titulable. Por último, la materia seca tiene un rango que va desde un 12 % hasta un 20 %.

TAREA 1.2.-ESPECIFICACIONES DE USUARIO Y OPERACIONALES.

[9] Peng, S.; Chen, C.; Shi, Z.; Wang, L. Amino Acid and Fatty Acid Composition of the Muscle Tissue of Yellowfin Tuna (Thunnus Albacares) and Bigeye Tuna (Thunnus Obesus). *Journal of Food and Nutrition Research*, Vol. 1, No. 4, 42-45, 2013.

[10] Karunarathna, K.; Attygalle, M. Nutritional evaluation in five species of tuna. *Vidyodaya journal of Science*. Vol. 15, No. 1 & 2, 7-16, 2010.

[11] García, J.C.; García, G.; Ciordia, M. El cultivo del kiwi. Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA), 2015.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACIÓN PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACIÓN CIENTÍFICA APLICADA Y LA TECNOLOGÍA
iasturias INNOVACIÓN

CAFÉ:

En el caso del café, se están desarrollando varias aplicaciones. Así, el punto del proceso productivo donde las aplicaciones desarrolladas podrán ser implementadas depende de la aplicación:

- % en Peso de Café Arábica/Robusta (Café verde): Esta aplicación NIR tiene su razón de ser en la zona de recepción de materia prima. En este lugar es donde llegan los camiones con los sacos de café verde. Generalmente, se realiza una inspección visual, se ve el tamaño del grano y se realiza una cata (el café se tuesta). Justo a la llegada del camión, cuando se realizan estas comprobaciones es el lugar idóneo para esta aplicación NIR. Se presentan dos posibilidades:
    - Escanear con el espectroscopio NIR el café directamente en la zona de recepción.
    - Llevarlo al laboratorio o zona habilitada de la empresa y realizar allí la medición NIR. En este caso, es factible llevar la muestra de café verde al laboratorio, ya que el tiempo no es crítico en esta parte del proceso. Por ejemplo, se podría realizar la medida NIR mientras se tuesta el café para realizar la cata, ya que el proceso de tueste lleva más tiempo que la realización de una medida NIR, y además debe realizarse en una tostadora pequeña para tal efecto que suele estar ubicada en el laboratorio.

  En las plantas productivas de café, se valora como más atractiva la opción del laboratorio, ya que generalmente la zona de recepción está en exteriores con cambios importantes de temperatura no solo diarios sino estacionales. Además, la necesidad del tueste del café para la cata obliga a llevar el café al laboratorio, que es un ambiente más controlado. Otro aspecto a tener en cuenta es que de esta forma las medidas NIR se realizarían en el mismo sitio que el resto de pruebas que se le hacen a la materia prima recepcionada.

  En las medidas NIR se deben utilizar muestras y técnicas de medida que sean lo más representativas posibles de la población. En el caso de esta aplicación de mezclas de Arábica/Robusta, al ser el café en grano y cada grano de una variedad, este aspecto es especialmente importante. Si el procedimiento de medida NIR solo es capaz de medir una pequeña superficie, solo aparecerán unos pocos granos de café (incluso puede darse el caso de solo 2-3 granos), con lo que la medida no será representativa de la muestra. La técnica de medida elegida deberá lograr una buena representatividad de la muestra. En principio para esta aplicación se está pensando en medidas a distancia o con el promediado de muchas medidas en contacto en diferentes puntos de la muestra. No serán necesarios ciclos de limpieza constantes, ya que el café verde no ensucia la sonda. Únicamente será necesario un mantenimiento general del dispositivo para mejorar su durabilidad.

- Parámetros relacionados con el grado de tostado del café (grano tostado): Los parámetros para los que se construirá una aplicación para su determinación a la salida de la tostadora son la densidad del grano, el color, la humedad y la pérdida de peso. El lugar donde se debe realizar la medida NIR para la determinación de estos parámetros es a la salida de la tostadora de café. El café sale de la tostadora a 200ºC y se enfría rápidamente con un aspirador que succiona el aire caliente proveniente de los granos de café tostado a la salida de la tostadora. Esto es necesario para evitar que el café se siga tostando. Los granos de café disminuyen su temperatura de 200ºC a 29ºC aproximadamente. Esta temperatura final no tiene que ser siempre la misma, depende en gran medida de las condiciones ambientales exteriores.

  La zona donde se debería incluir la medición NIR para esta aplicación sería justo después de la zona de enfriamiento del café salido de la tostadora. Normalmente, una vez enfriado, el café tostado es enviado a los silos de almacenamiento por medio de cintas transportadoras y aspiradores. Actualmente en este punto es donde se realizan los controles de color y densidad

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

que se realizan por un operario justo al lado de la línea productiva. También se mide la humedad del café, pero este es enviado al laboratorio para determinar la humedad, no se determina in situ y su medida tarda varios minutos, de forma que en ocasiones puede darse el caso de que se tuesten varias tandas de café antes de obtener el resultado de la primera de ellas.

Existen dos posibilidades de implementación de la medida NIR:

- Realizarla por un operario al lado de la línea de producción.
- Realizar la medida NIR automáticamente sobre la línea de producción, que sea una parte automatizada del proceso. La medida se realizará cuando el café tostado pase sobre la cinta transportadora, estando la sonda del equipo NIR sobre la cinta transportadora.

Como en el caso de la aplicación de las variedades Arábica/Robusta, en este caso también se está valorando realizar medidas a distancia o con el promediado de muchas medidas en contacto en diferentes puntos de la muestra, dado la heterogeneidad que presentan los granos de café. La limpieza no supone un problema en ningún caso, ya que la necesidad de limpieza es mínima si se realizan a distancia, y si se realizan en contacto tampoco será un problema porque el café en grano tostado ensucia poco. La diferencia con la aplicación anterior es que en este caso la aplicación debe tener en cuenta los tiempos del proceso del tostado del café. Cada tostada de café tarda algo más de 20 minutos, a los que hay que añadir un período de tiempo muerto entre tostadas. En estos rangos de tiempo, no va a haber problema ya que las medidas NIR se mueven entre los rangos de segundos/décimas de segundo. Estos tiempos comparados con los más de 20 minutos de cada tostada, no van a suponer un problema en la posible ralentización del proceso.

- Humedad del café crudo, grano tostado y molido: La determinación de la humedad en estos tres puntos del proceso productivo, es una necesidad transmitida por las empresas cafeteras de tener la humedad del café controlada en todo el proceso productivo. Estas zonas son especialmente importantes por las siguientes razones:
  - Café verde: Medir la humedad del café verde en recepción. Esto es importante porque una alta humedad es indeseable, ya que hace variar las condiciones de tostado y puede presentar más problemas de índole microbiológica.
  - Humedad del café salido de la tostadora: Esto se comentó en el punto anterior. Una alta humedad en esta zona conlleva una disminución en la calidad del café.
  - Humedad del café molido: Consiste en medir el café después de su molienda. Como en el caso anterior una alta humedad conlleva una disminución en la calidad del café. Pero en este caso es un punto más sensible, donde el café adquiere humedad mucho más rápido que el café en grano, ya que como es bien conocido, el café es muy higroscópico una vez molido.

Para el caso del café en grano tanto verde como tostado, se presentan los mismos aspectos que en las aplicaciones anteriores, heterogeneidad de muestras, medida a distancia o con el promediado de muchas medidas en contacto en diferentes puntos de la muestra, tiempos de proceso largos comparados con los tiempos necesarios para medidas NIR y facilidad de limpieza. Para el caso del café tostado molido, es algo diferente porque la heterogeneidad de muestra es un aspecto de menor importancia debido a la molienda del café. Y se necesita una pequeña limpieza, retirando el café de la sonda en caso de realizar las medidas en contacto.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

ATÚN:

La aplicación NIR del atún consiste en determinar el tipo de especie. El punto de la cadena de producción donde las empresas están interesadas en determinar la especie de atún es después de cocer el atún y justo antes de enlatar el atún. Cuando el atún este cocido y limpiado solo se puede identificar por ADN. En este punto del proceso, normalmente hay que esperar a que salgan los atunes del cocedero y se limpien, por lo que la acción de la espectroscopia NIR sería discontinua. Normalmente esta zona es una zona amplia, porque suele haber gente que está cortando y limpiando los atunes, por lo que necesita sitio para facilitar su movilidad. La instalación de un espectrofotómetro NIR no es un problema, podrá estar en la línea de producción como en un anexo a esta. El tiempo de escaneo es importante, pero más importante es lograr una correcta identificación, para lo que es necesario realizar varios escaneos en diferentes puntos de las muestras para lograr salvar las heterogeneidades de las muestras y alcanzar la representatividad suficiente en la media de las medidas, y salvar posibles problemas asociados a su geometría. Se deben realizar tomas de blanco (referencia), en principio, cada hora. Entre medidas, con una limpieza rápida (retirar restos) de la superficie de la sonda es suficiente.

KIWIS:

En el caso de los kiwis, se busca una aplicación para la cual mediante una medida NIR se logren 3 propiedades físicas simultáneamente. Existen dos posibles zonas donde pueden producirse las medidas NIR:

- En la plantación. En este caso, se debe poder medir en los kiwis que se encuentran en la planta, sin desprenderlos ni destruirlos. Permitiendo su futuro desarrollo en frutos aptos para su comercialización y evitar pérdidas de producto en análisis químicos.
- Postcosecha. Las medidas se producirán después de la cosecha del fruto.

En los dos casos será necesario un NIR que se pueda mover a las zonas donde van a producirse las medidas. En el caso de la plantación, deberá contar con una sonda con cables protegidos los suficientemente largos y que permitan la maniobrabilidad necesaria para realizar las medidas en el arbusto.

Respecto a los tiempos de proceso, en este caso no son limitantes para los dos casos expuestos antes, ya que no existe una cadena de producción, simplemente se tratará de medidas aleatorias para conocer la evolución de las frutas, no existiendo un intervalo temporal entre medidas. El operador tiene total libertad en este sentido. En este caso no son necesarios ciclos de limpieza constantes, ya que la medición sobre la superficie de los kiwis, sin destrucción que ensucie la sonda del dispositivo. Únicamente, será necesario una rápida limpieza de la superficie de la sonda en caso de adherirse pequeñas partículas si la sonda tiene un cristal protector que sirva de tope a la muestra, en caso contrario no es necesario limpiar entre medidas nunca. En caso de medir el espectralón (tomar el blanco), conviene realizar una limpieza por las razones expuestas anteriormente para el caso de haber cristal protector, y para todo tipo de sondas por pura protección del espectralón, para evitar que se ensucie. Los kiwis tienen forma oval, por lo que las sondas utilizadas deben ser capaces de amoldarse a la geometría de las frutas.

HITO 2.- DISEÑO EXPERIMENTO Y ADQUISICIÓN DE DATOS.

Tarea 2.1.-Diseño del protocolo experimental para asegurar la exactitud y la confiabilidad.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

REVISIÓN BIBLIOGRÁFICA:

CAFÉ:

Estimación de la composición de mezclas para Arábica/robusta:

Existen varios estudios que buscan relaciones entre la información espectroscópica de la región visible/NIR y la composición de mezclas de cafés Arábica/Robusta:

- (Esteban-Díez et al., 2004)[12] utiliza dos métodos de corrección de señal ortogonal (OSC y DOSC) sobre los espectros de 83 muestras de café de diferentes variedades y orígenes. Con este tratamiento de los espectros se busca eliminar la información no relacionada con la respuesta química específica de la cafeína. El café de variedad Robusta tiene más cafeína que el Arábica, por lo que esta señal es especialmente adecuada para discriminar entre ambos tipos de café. A parte de la cafeína, también se utilizaron los ácidos clorogénicos y acidez total en la clasificación de las variedades de café. Las muestras, que corresponden a 36 muestras de Arábiga y 47 de Robusta, fueron tostadas a diferentes condiciones de tostado. La utilización de la cafeína como respuesta química específica aporta los mejores resultados con un 100 % de aciertos en datos usados en la creación del modelo (clasificación) como en datos externos no usados en la creación del modelo (predicción).
- (Downey et al., 1994)[13] estudian la discriminación entre Arábica, Robusta y mezclas de los dos tipos de café mediante espectroscopia de reflectancia NIR. El estudio abarca tanto café en grano como molido. Para clasificar entre tipos de café entero Arábica y Robusta (tanto verde como tostado), se alcanza un 96.2 % de aciertos, pero cuando se incluyen mezclas de los dos tipos de café al 50 % de cada tipo como una categoría más, reduce el % de aciertos entre 82.9 y 87.6. Para el caso del café molido se alcanza un 83.02 % de aciertos, incluidos las mezclas.
- (Esteban-Díez et al., 2007)[14] utilizando del método de preprocesado de la corrección de señal ortogonal directa (DOSC) y el método de las funciones potenciales como técnica de modelado de clase. Utilizan 83 mezclas de café Arábica y Robusta puros, y 108 mezclas de ambos cafés que van desde un 0 % hasta un 60 % de café Robusta, añadiendo 3 categorías extra (5-20%, 25-40% y 45-60% de Café Robusta en mezcla). Se alcanzan unos buenos en la clasificación no solo de los tipos de café puros sino también en las mezclas de ambos. Para tres categorías (Arábica, Robusta y mezcla de ambos) se logra una correcta clasificación del 100% de las muestras en los datos usados en la creación del modelo y un 98.9 % para los datos de test (no usados en la creación del modelo). Para el caso de las cinco categorías (Arábica, Robusta y tres rangos de % en peso de la mezcla de ambos) se logra un 100 % de clasificación tanto en las muestras usadas para construir el modelo de clasificación como en las muestras de testeo del modelo.
- (Davieux et al, 2007)[15] desarrollaron un modelo para predecir la cantidad de café Arábica en mezclas de café Arábica/Robusta de café molido a cinco % de café arábica (Arábica puro, Robusta puro, y mezclas de Arábica/Robusta a 25%, 50% y 75% de café Arábica). Además, se realizaron tres tipos de tueste en función del tiempo que el café está en la tostadora: claro,

[12] Esteban-Díez, I.; González-Sáiz, J.M., Pizarro, C. An evaluation of ortogonal signal correction methods for the characterization of arabica and robusta coffee varieties by NIRS. *Analytica Chimica Acta* 514 (2004) 57-67.
[13] Downey, G.; Boussion, J.; Beauchêne, D. Authentification of whole and ground coffee beans by near infrared Reflectance spectroscopy. *Journal of Near Infrared Spectroscopy*. 2 (1994) 85-92.
[14] Esteban-Díez, I.; González-Sáiz, J.M.; Sáenz-González, C.; Pizarro, C. Coffee varietal differentiation based on near infrared spectroscopy. *Talanta* 71 (2007) 221-229.
[15] Davrieux, F.; Bastianelli, D.; Flori, A.; Manez, J.C.; Guyot, B. Influence of roasting level on near infrared spectroscopy prediction of Arabica content from ground coffee blends. *Near infrared spectroscopy : Proceedings of the 12th International Conference*. Auckland, New Zealand, 2005.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

medio y oscuro. Se logra un Error standard de predicción (SEP) de 4.21 %.

- (Downey et al., 1997)[16] utiliza análisis discriminante factorial y mínimos cuadrados parciales para clasificar 56 muestras de café tostado, molido y liofilizado de las variedades Arábica y Robusta. Ha parte de usar espectroscopia NIR usa espectroscopia MIR (infrarrojo medio). Para el caso NIR se logra un 100% de muestras correctamente clasificadas.
- (Correia et al., 2018)[17] prepararon mezclas de café molido Arábica con cáscaras/palos, maíz y café Robusta a tres niveles de tostado (claro, medio y oscuro). Las concentraciones de las impurezas van desde el 0% al 100%. Las muestras fueron medidas con un NIR portátil y se utilizaron las técnicas del análisis de componentes principales y de mínimos cuadrados parciales, con los pretratamientos spectrales previos, para construir modelos quimiométricos para clasificar y predecir % en peso de café Robusta y otros adulterantes. Utilizan 16 marcas diferentes de café, para el caso de adulteración con café Robusta, que sirven para crear 11 mezclas a diferentes concentraciones en peso de Arábica/Robusta (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 % en peso). Se crean modelos quimiométricos para todas las combinaciones posible de café Arábica y Robusta con los tuestes claro, medio y oscuro, en total nueve modelos diferentes. El modelo con mejores predicciones es el que se creó con datos de Arábica con tueste claro y Robusta con tueste claro con una raíz cuadrada del error cuadrático medio de predicción (RMSEP) de 2.8 % en peso. Los modelos quimiométricos se mueven en unos RMSEP entre 2.8 y 6.6 % en peso, siendo unos buenos resultados.
- (Pizarro et al., 2007)[18] usan la espectroscopia NIR para determinar el contenido de café Robusta en mezclas de café tostado Arábica/Robusta, que van desde un 0% hasta un 60% en peso de café robusta. En total son 191 muestras de café tostado, 83 muestras de café puro (Arábica y Robusta) y 108 mezclas de ambos cafés. El mejor modelo quimiométrico es el método PLS junto el pretratamiento denominado OWAVEC (método de corrección ortogonal wavelet), el cual tiene un RMSEP de 0.79% en peso de café robusta.

<u>Humedad del café crudo, grano tostado y molido:</u>

- (Tugnolo et al., 2019)[19] comparan dos espectrofotómetros NIR, uno de sobremesa y el otro portátil, ambos con un rango espectral de 960-1650 nm. Para café molido, determinan contenido en humedad, densidad compactada y granulometría del polvo. De café tostado en grano se realizaron medidas de humedad. A parte de los dos espectrofotómetros usan 3 tipos de blends (mezclas) diferentes. Para el espectrofotómetro de sobremesa se obtiene un RMSECV de 0.13 para el café tostado en grano y de 0.14 para el molido. Para el espectrofotómetro portátil se obtiene un RMSECV de 0.19 para el café tostado en grano y de 0.16 para el molido.
- (Morgano et al., 2008)[20] determinan la humedad en muestras de café crudo. Utilizan 157 espectros de reflectancia de muestras de café crudo con rango espectral de 1000-2222 nm. Proponen dos modelos quimiométricos consistentes en un método PLS y diferentes

[16] Downey, G.; Briandet, R.; Wilson, R.H.; Kemsley, E.K. Near- and mid-infrared spectroscopies in food authentication: Coffee varietal identification. *Journal of Agricultural and Food Chemistry*. 45 (1997) 4357-4361.
[17] Correia, R.M.; Tosato, F.; Domingos, E.; Rodrigues, R.R.T.; Aquino, L.F.M.; Filgueras, P.R.; Lacerda, V.; Romão, W. Portable near infrared spectroscopy applied to quality control of brazilian coffee. Talanta. 176 (2018) 59-68.
[18] Pizarro, C.; Esteban-Diez, I.; González-Sainz, J.M. Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*. 585 (2007) 266-276.
[19] Tugnolo, A.; Beghi, R.; Giovenzana, V.; Guidetti, R. Characterization of Green roasted beans, and ground coffee using near infrared spectroscopy: A comparison of two devices. *Journal of Near Infrared Spectroscopy*. 27 (2019) 1-12.
[20] Morgano, M.A.; Faria, C.G.; Ferrao, M.F.; Bragagnolo, N.; Ferreira, M.M.C. Determinação de umidade em café cru usando espectroscopia NIR e regressão multivariada. *Ciência e Tecnoligia de Alimentos*. 28 (2008) 12-17.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

pretratamientos de los espectros NIR. Para los errores RMSECV de validación cruzada se obtienen errores de 0.625 y 0.754 % (humedad en base seca, peso muestra desecada entre peso muestra, por 1000), algo diferentes, pero para la validación externa, RMSEP se obtienen valores similares de 0.550 y 0.546%.

- (Davrieus et al., 2008)[21] elaboran un estudio con la finalidad de controlar la calidad del café a la salida de las tostadoras, en concreto la humedad del café tostado en grano. Utilizan 284 muestras de Arábica y Robusta tostadas en condiciones industriales y en laboratorio. Las muestras fueron rehumedecidas para lograr un rango de humedad del 1-6 % de humedad. Las muestras fueron medidas tanto en continuo como en estático. Para la predicción se usa el método de mínimos cuadrados parciales (PLS). Se obtienen resultados similares para los dos dispositivos NIR usados (industrial y laboratorio), con errores standard de predicción (SEP) de 0.19% y 0.18 %.
- (Adnan et al., 2017)[22] evalúan la capacidad predictiva de la espectroscopia de infrarrojo cercano para la predicción no destructiva de la humedad en granos intactos de café verde de las variedades de café Arábica y Robusta. Para ello usan un espectrofotómetro NIR de transformada de Fourier con rango entre 1000-2500 nm y con un recipiente para muestras rotatorio. El mejor modelo quimiométrico logra un RMSEP de 0.57 %, pero logran un modelo con tan solo 7 longitudes de onda con un RMSEP de 0.77 %. Los modelos son válidos tanto para café Arábica como Robusta.

<u>Parámetros relacionados con el tostado del café:</u>

Para el caso de la humedad, las citas fueron presentadas anteriormente. Para las otras propiedades relativas al tostado:

- (Esteban-Díez et al., 2004)[23] predicen la acidez total, contenido en cafeína, ácidos clorogénicos y el color del tostado a partir de medidas NIR. Las medidas NIR se realizan con espectrómetro con detector de reflectancia, con rango 1100-2200 nm, y un módulo de transporte de muestra para tratar de salvar las heterogeneidades de las muestras. Se realizan varios modelos quimiométricos para el ajuste el método PLS y diferentes pretratamientos. El mejor modelo obtenido para la predicción del color del café tostado logra un 2.8 % y un 4.87 % para el % RMSECV y % RMSEP.
- (Pizarro et al., 2007)[24] utilizan la espectroscopia NIR para predecir el contenido en cafeína y el color de tostado en grano. Para ello se utilizan las variedades Arábica y Robusta, tanto puras como en mezcla. Se usa un espectrómetro en modo reflectancia con un rango de 1100-2500 nm, pero solo se usa el rango de 1100-2200, ya que el rango 2200-2500 nm presenta mucho ruido. En la predicción del color, se logra unos RMSECV y RMSEP de 1.41 % y 1.62 %, respectivamente,

[21] Davrieux, F.; Berkia, A.; Raverdy, J.L.; Excoffon, S. On-Line roasted coffee quality control using NIR Spectroscopy. *22nd International Conference on Coffee Science. ASIC*. Montpellier. (2008) 720-723.
[22] Adnan, A.; von Hörsten, D.; Pawelzik, E.; Mörlein, D. Rapid prediction of moisture content intact Green coffee beans using near infrared spectroscopy. Foods. 6 (2017).
[23] Esteban-Díez, I.; González-Sáiz, J.M.; Pizarro, C. Prediction of roasting color and other quality parameters of roasted coffee samples by near infrared spectroscopy. A feasibility study. *Journal of Near Infrared Spectroscopy*. 12 (2004) 287-297.
[24] Pizarro, C.; Esteban-Díez, I.; González-Sáiz, J.M.; Forina, M. Use of Near-Infrared Spectroscopy and Feature Selection Techniques for Predicting the Caffeine Content and Roasting color in Roasted Coffees. *Journal of Agricultural and Food Chemistry*. 55 (2007) 7477-7488.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

con un modelo PLS con eliminación iterativa con pasos (ISE-PLS).

- (Alessandrini et al., 2008)[25] utilizan la reflectancia difusa con rango 830-2500 nm para estimar la humedad, pérdida de peso, humedad y color a partir de granos de café verde y tostados a diferentes grados. Las muestras son tanto de las variedades Arábica y Robusta puras, como de mezclas industriales de ellas. Se desarrollan modelos quimiométricos para la pérdida de peso, humedad y densidad, consistentes en técnicas de ajuste PLS con diferentes pretratamientos, asimismo no usan todo el rango espectral del que dispone el espectrofotómetro, sino que se centran en regiones concretas. Los errores de validación cruzada (RMSECV) obtenidos son de 1.13 %, 0.02 %, 0.38 % para la pérdida de peso, densidad y humedad, respectivamente. Los errores de predicción (RMSEP) obtenidos son de 1.23 %, 0.02 %, 0.39 % para la pérdida de peso, densidad y humedad, respectivamente.
- (Bertone et al., 2016)[26] elaboran un modelo que determina simultáneamente el grado de tostado y la relación que existe entre el peso de Arábica/Robusta en las muestras. En total usan 130 mezclas de café Arábica/Robusta que van desde 0% hasta 100 % peso de café Arábica. Usan un método i-PLS (PLS iterativo para la elección de intervalos de longitudes de onda) que como pretratamiento tiene un Savitzky-Golay (primera derivada). Los modelos quimiométricos tienen un error de validación cruzada (RMSECV) es de 1.18 A.U. (unidades de absorbancia) y de 4.58 % para el color de tostado y del % de contenido en Arábica, respectivamente. Los errores de predicción (RMSEP) de 1.28 A.U. (absorbance units) y 4.34 % para el color y el contenido en Arábica, respectivamente.

ATÚN:

Para el caso de la identificación de especies de atún no existe ninguna bibliografía que relacione la clasificación de especies de atún con la espectroscopia NIR.

KIWIS:

En el caso de los kiwis se pueden encontrar varios artículos científicos que abordan el estudio de las tres propiedades físicas de interés mediante espectroscopia NIR.

- (Feng et al., 2011)[27] utilizan un espectrofotómetro NIR portátil para determinar el contenido de materia seca, los sólidos solubles y el tono de la pulpa en el caso de los kiwis de pulpa amarilla (Actinidia chinensis). El espectrofotómetro portátil (FQA-NIRGUN de Fantec) tiene un rango espectral de 600-1000 nm logra unos RMSEC de 0.6%, 0.9% y 1.4º para materia seca, solidos totales solubles y tonalidad de la pulpa, respectivamente.
- (McGlone et al., 2002)[28] investigaron la materia seca y sólidos solubles en kiwis de la variedad Actinidia deliciosa – Hayward mediante espectroscopia NIR. Para ello usaron un set de 360 kiwis recolectados en estado inmaduro. La mitad de los kiwis se analizan inmaduros y la otra mitad cuando alcanzan la madurez. El espectrofotómetro usado tiene un rango de 300-1140 nm y

[25] Alessandrini, L.; Romani, S.; Pinnavaia, G.; Dalla Rosa, M. Near infrared spectroscopy: An analytical tool to predict coffee roasting degree. *Analytica Chimica Acta*, 625, 95-102, 2008.
[26] Bertone, E.; Venturello, A.; Giraudo, A.; Giraudo, A.; Pelligrino, G.; Geobaldo, F. Simultaneous determination by NIR spectroscopy of the roasting degree and Arabica/Robusta ratio in roasted and ground cofee. Food Control. 59 (2016) 683-689.
[27] Feng, J.; McGlone, A.V.; Currie, M.; Clark, C.J.; Jordan, B.R. Assessment of yellow-fleshed kiwifruit (Actinidia chinensis 'Hort16A') quality in pre- and post-harvest conditions using a portable near-infrared spectrometer. *Hortsicence*. 46 (2011) 57-63.
[28] McGlone, V.A.; Jordan, R.B.; Seelye, R.; Martinsen, P.J. Comparing density and NIR methods for measurement of Kiwifruit dry matter and soluble solids content. *Postharvest Biology and Technology*. 26 (2002) 191-198.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

logra unas predicciones de alrededor del 0.5% o incluso menos para las dos propiedades físicas.

- (McGlonw et al., 2007)[29] utilizan un espectrofotómetro de rango 300-1140 nm para predecir materia seca, sólidos totales solubles y color de la pulpa. Para los casos de la materia seca y los sólidos solubles totales se utiliza el rango espectral entre 800-1000 nm. Se consiguen unos RMSEP de 0.4% y 0.71% para materia seca y sólidos solubles totales, respectivamente, para frutas recién cosechadas. Para frutas almacenadas por 12 semanas a 0ºC, se logran unos RMSEP de 0.24% y 0.31% para materia seca y sólidos solubles totales, respectivamente.
- (Schaare et al., 2000)[30] utilizan tres modos de medida de infrarrojo cercano (reflectancia, interactancia y transmisión) con la finalidad de conocer sus capacidades predictivas para estimar contenido en sólidos solubles totales, densidad y color de la pulpa para kiwis de pulpa amarilla de la variedad Actinidia chinensis. El modo que aporta los mejores resultados es la interactancia. El modo de reflectacia utiliza un rango de longitudes de onda de 450-1100 nm y tiene un error de predicción standard (SEP) de 1.18 ºBrix en la predicción de la cantidad de sólidos solubles.
- (McGlone et al., 1998)[31] prueban la validez de la tecnología NIR para estimar la madurez postcosecha y propiedades fisicoquímicas de la variedad Actinidia deliciosa – Hayward. Se utilizan medidas de interactancia con un rango espectral de 800-1100 nm, con un error RMSEP de 0.42% para la materia seca y 0.39 ºBrix para los sólidos solubles.
- (Lü et al., 2010)[32] utiliza el método "Synergy interval partial least square" (siPLS) para elegir las regiones de longitudes de onda características para predecir la materia seca utilizando infrarrojo cercano. Para el estudio se han usado cuatro conjuntos de datos: Espectros NIR y análisis de materia seca de kiwis inmaduros, espectros de kiwis inmaduros y materia seca de kiwis maduros, espectros NIR y análisis de materia seca para kiwis maduros, y todos los casos anteriores juntos. Los resultados muestran el cambio de las regiones de longitudes de onda características fueron causadas por la conversión química de los compuestos orgánicos incluidos en la materia seca en diferentes periodos de almacenamiento. Los diferentes modelos construidos tienen unos RMSECV (validación cruzada) entre 0.39-0.48% y de RMSEP (predicción) de 0.41-0.53%.
- (Slaugther et al., 1998)[33] determinan la calidad interna (contenido de fructosa, glucosa, contenido en sólidos solubles y materia seca) de los kiwis mediante técnicas espectroscópicas de infrarrojo cercano. Usando un rango espectral de 400-1100 nm, logran predecir el contenido en sólidos solubles y materia seca con errores standard de calibración de 0.78 ºBrix y 0.61%, respectivamente.
- (Lee et al., 2012)[34] tratan de establecer un standard de clasificación de kiwi maduro con las muestras de tres haciendas. Para ello miden firmeza de la pulpa, contenido en sólidos solubles y

[29] McGlone, V.A.; Clark, C.J.; Jordan, R.B. Comparing density and VNIR methods for predicting quality parameters of yellow-fleshed kiwifruit (Actinidia chinensis). *Postharvest Biology and Technology*. 46 (2007) 1-9.
[30] Schaare, P.N.; Fraser, D.G. Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (Actinidia chinensis). *Postharvest Biology and Tecnology*. 20 (2000) 175-184.
[31] McGlone, V.A.; Kawano, S. firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology*. 13 (1998) 131-141.
[32] Lü, Q.; Tang, M.; Cai, J.; Lu, H. Long-term prediction of Zhonghua kiwifruit dry matter by near infrared spectroscopy. *ScienceAsia*. 36 (2010) 210-215.
[33] Slaughter, D.C.; Crisosto, C.H. Nondestructive internal quality assessment of kiwifruit using near-infrared spectroscopy. *Seminars in Food Analysis*. 3 (1998) 131-140.
[34] Lee, J.S.; Kim, S.C.; Seong, K.C.; Kim, C.H.; Um, Y.C.; Lee, S.K. Quality prediction of kiwifruit based on near infrared spectroscopy. *Korean journal of horticultural science and technology*. 30 (2012) 709-717.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

acidez. Junto con estas propiedades físicas se miden los espectros NIR (408-2492 nm) asociados a las muestras analizadas y con toda esta información se construyen modelos quimiométricos mediante un método de mínimos cuadrados parciales modificado. Se lograron SEP de 0.49 ºBrix y 0.28% para contenido en sólidos solubles y acidez, respectivamente.

- (Qiang et al. 2010)[35] usan varias técnicas para buscar longitudes de onda más eficientes en la predicción de materia seca. Entre las técnicas usadas están la técnica clásica de los mínimos cuadrados parciales (PLS) y otras más modernas como "synergy interval PLS" (siPLS) and siPLS junto caon algoritmo genético (GA-siPLS). El intervalo inicial de búsqueda de las mejores longitudes de onda es 1000-2500 nm. El método GA-siPLS tiene los mejores resultados con un RMSEP de 0.53%.

FUENTES DE VARIABILIDAD:

CAFÉ:

El café tiene varias fuentes de variabilidad como pueden ser su origen, época de cosecha y el tipo de café (Arábica/Robusta). Después de hablar con empresas del sector, se conoce que estos tres factores son altamente interdependientes. Las empresas compran el café en función del lugar del mundo donde se produzca la maduración, se pueda recolectar y exportar. Generalmente, las empresas tienen un sistema de compras anual cíclico siguiendo las cosechas alrededor del mundo, y de esta forma obtener el café más reciente. Así, los factores de origen y época de cosecha están muy fuertemente interconectados, ya que cuando se cosecha en un lugar suele ser de forma casi simultánea, en la misma época. El tercer factor, el tipo de café (Arábica/Robusta), también está relacionado con los dos anteriores. Las variedades de Arábica y Robusta no están distribuidas homogéneamente por el mundo, sino que se dan mejor en unas zonas que en otras. Se pueden encontrar los tres casos, países con producción mayoritariamente de la variedad Arábica, de Robusta y donde se producen cantidades importantes de las dos variedades. Por eso las muestras de arábica y robusta que se obtendrán serán dependientes del lugar de cosecha (como se comentó en la tarea 1.1). Otras fuentes de variabilidad son las condiciones de tostado. Las condiciones de tostado vendrán dadas por el grado, tipo de tostado y el tipo de café que se busca. El origen, tipo de café y humedad también condicionan el tostado. Para cubrir esta variabilidad se utilizarán diferentes tipos de café y diferentes tipos de tueste. Para ello se está valorando comprar café con diferentes grados de tueste o tostarlo nosotros mismos para lograr puntos precisos en el grado de tostado.

ATÚN:

En un principio se pensaba que la época de captura sería la fuente de variabilidad más importante. Pero no hay atún todo el año, la temporada del atún está restringida en el tiempo. Esta temporada va desde marzo a junio. Por otro lado, estudiando las piezas y hablando con las conserveras, se llegó a la conclusión que la más importante fuente de variabilidad va a ser la propia pieza de atún. Así, dependiendo donde se realice la medida se obtendrán espectros diferentes, por la propia distribución bioquímica del pez. No es lo mismo realizar una medida cerca de la superficie (cerca de la piel), que en el interior del pez.

KIWI:

Para el caso de los kiwis, existen dos fuentes de variabilidad, la zona de cultivo y el estado de maduración de la fruta. En este caso, los kiwis deben ser cultivados en Asturias, por lo que su época de crecimiento, recolección y postcosecha quedan reducidos en el tiempo. En Asturias, la especie de kiwi

[35] Qiang, L. Mingjie, T.; Jianrong, C.; Huazhu, L.; Chaitep, S. Selection of efficient wavelengths in NIR spectrum for determination of dry matter in kiwi fruit. *Maejo International Journal of Science and Technology*. 4 (2010) 113-124.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

más cultivada es la *Actinidia Deliciosa (kiwi verde), en concreto la variedad Hayward. Por esta razón, este estudio se centrará en esta variedad. Respecto a la maduración del kiwi existen dos fases claramente diferenciadas, antes de la cosecha y después de la cosecha. Se recolectarán suficientes kiwis de las dos épocas, siempre con la finalidad de tenerlas suficientemente cubiertas, con muestras distribuidas temporalmente a lo largo de ellas.*

ASPECTOS RELACIONADOS CON LA ESPECTROSCOPÍA NIR:

El mayor o menor número de escaneos viene determinado por el ruido presente en el espectrograma. A partir de un número de escaneos se puede considerar que el espectrograma no tiene o tiene muy poco ruido, por lo que no tiene sentido generar espectrogramas utilizando más escaneos de los necesarios. Por eso, se elegirá un número de escaneos lo suficientemente alto, para eliminar el ruido, para generar el modelo quimiométrico. Con el modelo quimiométrico construido se bajará el número de escaneos en las medidas hasta que las predicciones del modelo quimométrico empeoren debido al ruido generado por el descenso en el número de escaneos. Esta forma de actuar se aplicará a todas las aplicaciones quimiométricas desarrolladas en este proyecto.

En este proyecto se usarán dos espectrofotómetros, el Labspec 4 de ASD y un espectrofotómetro desarrollado por el centro tecnológico. Estos espectrofotómetros tienen las siguientes especificaciones:

- "Labspec 4 Standars-Res" de ASD: Este espectrofotómetro tiene un rango espectral que va de 350 nm a 2500 nm que cubren tres espectrómetros, un espectrómetro VNIR (con rango de 350-1000 nm) y dos SWIR (con rangos de 1001-1800 nm y 1801-2500 nm), con un total de 2151 canales equiespaciados en todo el rango espectral. La resolución espectral es de 3 nm en la zona VNIR y de 10 nm en la SWIR. Utiliza una sonda de reflectancia con luz halógena incorporada.
- Espectrofotómetro propio: El espectrofotómetro tiene un rango de 350-2100 nm que cubren dos espectrómetros, un VNIR (con rango 350-1000 nm) y un SWIR (con rango 900-2100 nm). Este espectrómetro tiene 2306 canales no equiespaciados, de los cuales 2048 corresponden al VNIR y 256 al SWIR. La resolución espectral es de 1.5 nm en la zona VNIR y de 7.6 nm en la SWIR. Utiliza una sonda de reflectancia con luz halógena incorporada.

CAFÉ:

Para tratar de paliar la heterogeneidad de las muestras en el caso del café, se abordarán las medidas de dos maneras diferentes en función del espectrofotómetro utilizado:

- Labspec 4: La sonda de este espectrómetro es de contacto con lo que para paliar la heterogeneidad de las muestras se deben realizar varias medidas en diferentes puntos de las muestras (10 medidas) para tener un espectro promedio de la muestra. En un principio se planificaron 5 medidas por muestra, pero se luego se decidió realizar 10 con el fin de lograr una mejor representatividad de la muestra. Estos 10 espectros se promediarán para obtener un espectro promedio representativo de la muestra
- Espectroscopio propio: La sonda de este espectrofotómetro tiene un módulo removible que al sacarse permite realizar medidas a distancia. Esta posibilidad se utilizará junto a un dispositivo que proporcione un movimiento circular de la muestra. Así, la lámpara y la sonda del espectrómetro que estarán situadas sobre el dispositivo giratorio, aprovecharán el movimiento circular para obtener un espectro representativo de la superficie de giro de la muestra. Además, conviene realizar varias medidas (3 por muestra), sacando la muestra de café del soporte de las muestras y volviéndola a colocar, para evitar posibles heterogeneidades propias de la colocación

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

de las muestras. La distancia a la muestra desde la sonda será la que proporcione la distancia focal óptima (proporcionada por el fabricante).

En el año 2019 se planteó la elección entre tostar el café (desde verde) o usar café tostado comercial (con diferentes grados de tostado). Al final se ha optado por usar café tostado comercial. Las razones son:

- Dificultad para tostar café sin una tostadora de café. Se han valorado opciones como tostarlo en un horno o en una sartén, pero el café tostado obtenido es muy desigual, sin lograr un mínimo grado de homogeneización requerida.
- Para todos los parámetros se puede usar café comercial sin problemas, incluso para el caso de la pérdida de peso, como se explicará a continuación. Para este caso es necesario saber el peso inicial del café verde antes de entrar a la tostadora, el peso inicial del café depende en su origen y sobre todo la cantidad de agua que contenga el café en su composición. Así, cuando un café verde lleva tiempo almacenado este suele tener una humedad mayor. Analizando la situación y hablando con las empresas productoras de café, se ha llegado a la conclusión que estas variaciones de humedad en el café inicial lo que provocan es que el café este algo más/menos tostado y tenga más/menos humedad al final (generalmente en la parte final del tostado se le añade vapor de agua al café), que el conocimiento de esta condición inicial es interesante para optimizar curvas de tostado de café, más que para para determinar la calidad del café. Como ejemplo de la problemática, cuando dos muestras de café verde con humedades determinadas y diferentes son tostadas con sus correspondientes curvas de tostada para dar un café similar, entonces el café al tener similares propiedades (sobre todo humedad) debería tener la misma pérdida de peso ya que la espectroscopia NIR se basa en la información aportada por la medida sobre la muestra, y esta es similar. Por estas razones se considera mejor usar un valor promedio para el peso del café inicial para todos los cafés, que sirva como referencia.
- Existen una gran variedad de cafés comerciales con diferentes grados de tostado, orígenes geográficos, orígenes botánicos, etc.

ATÚN:

En el caso del atún, las medidas serán en contacto. Dada la heterogeneidad de las muestras (por ejemplo, dependiendo de donde se tome la medida en la muestra, puede estar en diferente profundidad del pez), se deberán hacer múltiples medidas (5) en diferentes zonas (equidistantes unas de otras) para obtener un promedio representativo de la muestra. En este caso, se estima (por lo hablado con las empresas) que con medidas de menos de 1-2 segundos se amoldaría perfectamente a los ritmos de producción de las empresas. Se cree que se podrán lograr estos tiempos e incluso menores.

Finalmente se ha optado por utilizar producto acabado dado los problemas de conseguir atún de las empresas. En concreto, se han adquirido latas y botes de atún al natural de tres tipos:

- Bonito del norte. Científicamente conocido como *"Thunnus Alalunga"*. Es el que tiene un mayor valor.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

- Atún claro. Aunque solo la especie Yellowfin (*"Thunnus Albacanes"*) es atún claro, también se puede comercializar la especie bigeye *("Thunnus Obesus"*) bajo este nombre. Así, dos especies se pueden encontrar bajo este nombre.
- Atún. Lo más habitual es encontrar la especie Skipjack ("Katsuwomus Pelanis") bajo el nombre de atún a secas. Este producto es el que tiene menor precio.

KIWI:

En el caso del kiwi se realizarán medidas de contacto para los dos espectrofotómetros. En concreto se realizarán dos tipos de medidas:

- Kiwi entero: En este caso, los kiwis se miden a temperatura ambiente del laboratorio (aproximadamente 15-20ºC) y a baja temperatura (aproximadamente 4ºC). Para las medidas a baja temperatura, los kiwis se sacan de la cámara frigorífica (aprox. 4ºC) y se transportan al laboratorio en una bolsa para conservar el frío con hielo para refrigerar. Para los dos casos, se realizarán dos medidas sobre la zona ecuatorial en lados opuestos del kiwi. Si el kiwi no es totalmente circular (forma elíptica del ecuador), se realizará perpendicular sobre el semieje menor.
- Kiwi interior: El kiwi se cortará por su ecuador. Se realizarán dos medidas, una sobre cada mitad del kiwi abierto. La sonda se colocará sobre el borde de la superficie del kiwi tal como se indica en la figura 1. Si el kiwi no es totalmente circular (forma elíptica del ecuador), se realizará sobre el semieje menor. La temperatura de medida será la del laboratorio.

Se realizan dos medidas de espectroscopia NIR a diferentes temperaturas para evaluar posibles efectos de la temperatura. Del interior solo se realizan medidas a temperatura del laboratorio.

En las figuras 2.1.1 y 2.1.2, se pueden ver las zonas de medida para kiwis enteros para alta y baja temperatura (izquierda) y el interior (derecha). Las formas de anillo y rectángulo indican las zonas de medida para las sondas de los dos espectrofotómetros. El grosor de la zona azul de la circunferencia y del rectángulo son indicativos de la zona de la carcasa de la sonda donde no hay medida. La zona interior del anillo y el rectángulo, son las zonas de medida. En el caso del espectrofotómetro Labspec 4 se puede ver que la zona circular de medida está en el centro del kiwi para las medidas en el exterior, mientras que en el espectrofotómetro propio está desplazada hacia un polo del kiwi. Esto se debe a que en el espectrofotómetro Labspec 4 la luz de la sonda está centrada en el círculo azul, mientras que en el propio está hacia un extremo del rectángulo. Por esto es necesario colocar el extremo del rectángulo donde hay más luz en el ecuador del kiwi, para maximizar la cantidad de luz reflejada. En los interiores del kiwi, las zonas de medida se colocan sobre la pulpa evitando la columela, ya que es la pulpa la que está más relacionada con las propiedades estudiadas.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias Innovación

Figura 2.1.1.- Zona de medida de espectroscopia NIR sobre la superficie de un kiwi para el espectrofotómetro – Labspec 4 (ASD).

Figura 2.1.2.- Zona de medida de espectroscopia NIR sobre la superficie de un kiwi para el espectrofotómetro propio.

A lo largo del año 2019 se han analizado los resultados parciales del interior de los kiwis (kiwis partidos a la mitad). Los resultados, muestran resultados similares a los obtenidos con el kiwi entero. La idea de estas medidas del interior del kiwi tiene interés en el caso de lograrse claramente mejores calibraciones que para el caso del kiwi entero. No tiene interés destruir kiwis para lograr un resultado similar al obtenido sin destruirse. Por lo que a partir de ahora se medirán solo kiwis enteros a alta (aproximadamente 15-20ºC) y baja (aproximadamente 4ºC) con las cuales se realizará un modelo quimiométrico (como era originalmente el proyecto, medidas en superficie del kiwi sin destrucción de la fruta).

Respecto al número de escaneos realizados por el espectrómetro se ha tomado la decisión de usar 50 escaneos por medida para todas las aplicaciones.

En base a los cálculos y modelos realizados (todas las matrices alimentarias), se ha decidido seguir el trabajo solo con un espectrofotómetro (como era originalmente el proyecto), el modelo Labspec 4 de ASD. Las razones son los resultados inferiores que presenta el otro espectrómetro respecto al Labspec 4 (más de un diez por ciento peores). Teniendo en cuenta este hecho, junto con que el espectrómetro Labspec 4 tiene mayor rango espectral que el otro espectrómetro, en un futuro se puede usar el espectrómetro Labspec 4 como base para transferir las calibraciones realizadas hacia el otro espectrómetro, no siendo necesario realizar las calibraciones a la vez en los dos espectrómetros. Así, el elemento clave es el espectrómetro Labspec 4 y las calibraciones realizadas con él. Una vez realizadas no solo se podrán transferir al otro espectrómetro, sino también a otros espectrómetros de otras

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

empresas. Servirá como base de trabajo.

Tarea 2.2.- Implementación de diseño experimental de la adquisición de espectros.

A partir del diseño experimental se generaron las muestras necesarias que cumplan los requisitos especificados en la tarea 2.1. Cada muestra tiene asociado una serie de espectros (número especificado para cada aplicación) que se han promediado para obtener un espectro representativo de la muestra. Esto se ha hecho para todos los parámetros que se estimarán a partir de las calibraciones NIR.

KIWIS.

A lo largo del proyecto se han generado muestras para el caso de todas las aplicaciones de los kiwis. Así, durante tres años se generaron muestras de kiwis desde antes de alcanzar la maduración hasta sobrepasar la maduración ideal para consumo. A parte de las medidas de los espectros, se aplicaron los métodos analíticos correspondientes. Así para el caso de los kiwis, se han determinado:

- Grados Brix.
- Residuo Seco.
- Acidez titulable.

En general, la recolección del kiwi asturiano comienza entre la primera y la segunda semana de noviembre. Solo hay una época de cosecha en todo el año. El primer año de proyecto se utilizaron kiwis aptos para el consumo. En el segundo y tercer años se amplió el rango de kiwis desde antes de la cosecha (inmaduros) hasta sobrepasar la maduración ideal para consumo. Para ello se recogieron kiwis de la plantación desde dos semanas antes de la fecha de inicio de recolección de las bayas. En el inicio de recolección, se recogieron más kiwis para ir utiliándolos a lo largo de los días para su análisis obteniendo diferentes grados de maduración. También se adquirió kiwi asturiano en el supermercado para su análisis. Los kiwis según llegaron a las instalaciones de la empresa, fueron almacenados al menos 24 horas a 4 ºC en cámara frigorífica, con la finalidad de que se lograra la temperatura inferior para las medidas NIR de baja temperatura. Antes de realizar las medidas a alta temperatura, se dejaron atemperando en el laboratorio aportando aire con un ventilador, de forma que mediante convección forzada se logre subir y estabilizar su temperatura a la temperatura ambiente del laboratorio (alrededor de 20 ºC). Este proceso dura alrededor de 2-3 horas. De esta forma, para los casos en que se realizaron medidas NIR a temperatura alta y baja (la gran mayoría) para un mismo conjunto de kiwis, primero se medirán a baja temperatura (4ºC) y después a alta temperatura (aproximadamente 20 ºC).

Para realizar las medidas NIR de un conjunto de kiwis, se recogen los kiwis a medir dentro de la cámara frigorífica utilizando una bolsa térmica con hielo para mantener la temperatura. Para evitar el contacto de los kiwis con el hielo se introduce un separador para crear dos compartimentos en la bolsa, pero dejando espacio para el tránsito del aire dentro de la bolsa térmica.

Para tomar espectros con el espectrofotómetro NIR (Labspec 4), el protocolo de medidas utilizado hasta la fecha comienza con la comprobación de que todas las conexiones del espectrómetro están correctamente conectadas (tanto eléctricas como fibra óptica). Esta comprobación se implementó

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

dentro del protocolo del sistema NIR, ya que este se usa en la empresa en multitud de aplicaciones de diversa índole que conllevan cambios de sonda y desplazamientos fuera del laboratorio. En caso de que el espectrómetro no se encuentre con la configuración adecuada para medir kiwis, se deben realizar los cambios necesarios en las conexiones. Después de esto, primero se enciende el espectrómetro y luego la luz de su lámpara halógena. Se deja calentar la luz de la lámpara alrededor de 15-20 minutos. Después de este tiempo se enciende el ordenador y se entra en el programa de control del espectrómetro Indico Pro. Se toma la referencia (Spectralon) y ya se pueden realizar medidas NIR. Por cada kiwi, tanto a alta temperatura como a baja temperatura se realizan dos espectros en lados opuestos de la zona ecuatorial del kiwi (figura 2.2.1). Se tratará de tomar las medidas NIR en las zonas más planas de la parte ecuatorial del kiwi. Estas dos medidas de ambos lados de la superficie del kiwi se promedian, obteniéndose un espectro más representativo del kiwi.

Figura 2.2.1.- Medida NIR en la zona ecuatorial del kiwi.

Las medidas de baja y alta temperatura tienen sus propias carpetas y archivos de adquisición propios. Por eso ha sido necesario cambiar de archivos cuando se cambia de temperatura. Esto conlleva salir del archivo de adquisición de baja temperatura, entrar en el de alta temperatura y medir de nuevo la referencia. Además, se ha medido la referencia cada media hora aproximadamente o en caso de observarse que los espectros obtenidos presentan discontinuidades entre sensores. Esto se debe a derivas en la temperatura de la lámpara, especialmente visibles en temperaturas ambiente frías. En el caso de los kiwis (noviembre) y que se empiezan a medir a baja temperatura a primera hora de la mañana, hace que este problema haya sido habitual y que se tuvieran que realizar blancos varias veces, ya que las medidas a baja temperatura se han realizado en una zona de la empresa sin calefacción, para tratar de mantener la baja temperatura de los kiwis cuando se miden a baja temperatura.

También, hasta el segundo año, los kiwis han sido medidos con el espectrofotómetro propio. Los pasos seguidos en las medidas han sido exactamente iguales a los del Labspec 4. También se han realizado

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

medidas al interior de los kiwis, conforme a la tarea 2.1. Las medidas del interior de los kiwis son realizadas únicamente a alta temperatura (aproximadamente 20 ºC), ya que implican cortar el kiwi ecuatorialmente, lo que imposibilitaría posteriormente medir el kiwi entero a otra temperatura. Las medidas NIR en el interior de kiwis son las últimas medidas a realizar antes de aplicar el método analítico, porque conllevan la destrucción del kiwi. En las medidas del interior del kiwi, también se realizan dos medidas a ambas mitades del kiwi, siguiendo los pasos comentados en la tarea 2.1. De estos dos espectros también se obtiene un espectro promedio.

De esta forma general los pasos seguidos en la toma de espectros NIR de kiwis son los siguientes:

- Puesta a punto de los equipos NIR para las medidas.
- Medidas de kiwis enteros a baja temperatura con los dos espectrómetros.
- Lograr que los kiwis alcancen la temperatura ambiente.
- Medidas de kiwis enteros a alta temperatura con los dos espectrómetros.
- Medidas de interior de kiwis con los dos espectrómetros.
- Realización de los métodos analítico.

Una vez se han realizado todas las medidas referentes a la espectroscopia se procede con los métodos analíticos que se deben aplicar a los kiwis.

Extracto Seco.

El primer método analítico que se le ha realizado a los kiwis es la determinación de extracto seco. Lo primero que se debe hacer en este método es poner en una estufa a más de cien grados, durante un par de horas, los recipientes de los kiwis para eliminar toda humedad de su superficie. Pasado este tiempo se sacan de la estufa y se ponen en un desecador hasta su uso, para evitar que cojan humedad de nuevo. Los kiwis se han cortado a la mitad por la zona ecuatorial (puede estar ya cortado debido a las medidas NIR del interior de los kiwis) y de cada una de las mitades y se ha cortado una sección ecuatorial (dos en total) que debe ser muy fina (3 mm como mucho). Se toma un recipiente del desecador y se pesa en una balanza, se anota su peso. Entonces las dos secciones del kiwi se depositan en este recipiente y se pesa todo, anotando de nuevo su peso. El recipiente con los kiwis se introduce en una estufa a 65 ºC durante 24 horas. Pasado este tiempo se saca de la estufa y se introduce en el desecador durante 20 minutos, con la finalidad de que el recipiente se enfríe y las corrientes de convección dentro de la balanza no desestabilicen las medidas. Una vez frío, se pesa de nuevo y se anota el valor. Con los tres valores conseguidos se determina el valor del residuo seco mediante la siguiente ecuación:

$$\%RS = \frac{S - C}{H - C} \times 100$$

Donde %RS son los % de residuo seco, C es el peso del recipiente vacío, H es el peso del recipiente con las dos secciones de kiwi húmedas y S es el recipiente con las dos secciones de kiwi secas.

Grados Brix.

Para el caso de los grados Brix, antes de empezar fue necesario limpiar bien con agua destilada, y

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

después secar bien la superficie del prisma donde se depositan las gotas de jugo de kiwi, con el fin de evitar impurezas. Seguidamente se calibra el refractómetro con agua ultrapura, y se seca convenientemente. En este punto se realizaron las medidas. Para ello se ha colocado una gota de jugo de kiwi sobre la superficie del prisma y se pulsó el botón de medida, después de unos segundos se obtiene el valor de los grados Brix. Después de cada medida es necesario limpiar bien con agua destilada y secar la superficie del prisma.

Acidez Titulable.

Para determinar la acidez titulable se han usado 5 ml de jugo de kiwi. Para titularlos se ha usado NaOH 0.1 N. El procedimiento comienza pesando un vaso de precipitados pequeño y se anotando su peso. Luego, se depositan 5 ml de jugo de kiwi en el vaso de precipitados, se vuelve a pesar y se anota su valor. El peso del jugo de kiwi es el valor del peso del vaso con jugo menos el peso del vaso sin jugo. Se llena una bureta (que se coloca en un pie) con el NaOH, enrasándola adecuadamente. Se introduce un peachímetro digital para controlar el pH en todo momento durante la valoración. Se echa sosa desde la bureta hasta alcanzar pH 8.2. en este punto se para la valoración. Se anota la cantidad de sosa gastada. Este proceso se debe hacer despacio y con sumo cuidado, ya que las medidas del peachímetro son muy oscilantes, ya que la mezcla entre la sosa y el jugo no es instantánea. Así, se debe agitar en círculos el vaso de precipitados cada vez que se añada un poco de sosa para asegurarse una correcta mezcla.

$$AT = \frac{(Na \times 0.0064 \times 100)}{Pj}$$

Donde *Na* son los ml de NaOH 0.1 N y *Pj* es el peso en gramos de jugo analizado (alrededor de 5 g.)

A modo de resumen, en la siguiente tabla 2.2.1 se pueden ver las muestras generadas y espectros tomados para realizar los modelos de los parámetros de los kiwis.

Tabla 2.2.1.-Muestras usadas para la construcción de modelos matemáticos y espectros para los kiwis.

	Nº Muestras	Nº Espectros
Grados Brix	241	864
Residuo Seco	190	750
Acidez Titulable	154	578

Decir que se presentan los datos referentes al espectrómetro Labspec 4, ya que va a ser el espectrómetro para el que se van a seguir realizando medidas para el resto del proyecto, una vez descartado el otro espectrómetro como se comenta en la tarea 2.1. Tampoco se presentan los datos de las medidas del interior de los kiwis, porque como también se ha visto en la tarea 2.1 es otra línea

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias Innovación

descartada.

CAFÉ.

En los años 2019 y 2020, se ha realizado la generación y medida de muestras de las aplicaciones relativas al café. Se han realizado tanto medidas NIR como los métodos analíticos para la creación de los modelos de predicción de los siguientes parámetros:

- Relación Arábica/Robusta.
- Humedad del café verde.
- Humedad del café tostado entero.
- Humedad del café tostado molido.
- Color del café tostado entero.
- Densidad del café tostado entero.
- Pérdida de peso del café tostado entero.

En este caso (café), indicar que antes de generar muestras ha sido necesario una puesta a punto más larga y laboriosa que en los casos del kiwi y el atún. Ya que, en este caso, no solo hay más parámetros, sino que para estos parámetros ha sido necesario un mayor esfuerzo y tiempo (con el consiguiente problema de que estas labores no generan muestras) con el fin de poner a punto los métodos tanto espectrométricos como analíticos. Como consecuencia de esto, han sido necesarias correcciones en algunos métodos con las consiguientes repeticiones en la generación de muestras. Un caso claro de este aspecto han sido los problemas surgidos en la generación de muestras de distinta humedad, a la que se aportaron soluciones que se comentarán un poco más adelante en esta misma tarea.

En este caso, las formas de medida realizadas con el espectrómetro (Labspec 4) se han realizado de dos formas en función del tipo de muestras, café entero (tostado o verde) o café molido. Esto es debido a que la diferente forma física en que se presenta el café determina la mejor forma de medida NIR.

Como bien se comentó en la tarea 2.1, con el espectrómetro principal (Labspec 4) se realizaron 10 medidas (espectros) utilizando el recipiente creado para tal fin. Esta es la razón de tan gran diferencia entre el número de muestras y número de espectros, que en este caso es incluso mucho mayor que en el caso de los kiwis. Para el caso del café entero se llena el recipiente completamente, y en el caso del café molido se llena mediado. En este caso, para lograr una mayor representatividad se ha colocado menos muestra en el recipiente que en el caso del café entero. Los pesos del café en el recipiente de medidas NIR para los diferentes tipos de café son 13 g para el café verde (recipiente lleno), 7 g para el café entero tostado (recipiente lleno) y entre 3 y 4 g para el café tostado molido (recipiente mediado). Comentar también, que todas las medidas NIR se han realizado en laboratorio a aproximadamente 20 ºC.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 2.2.2.- Aquí se puede ver el recipiente utilizado para la medida de café y como se toma un espectro NIR.

El protocolo general de medida con la espectroscopia NIR de las muestras de café es similar al de los kiwis, con la salvedad del proceso de la toma de espectros que, como se ha visto, las medidas se realizan de forma diferente. Así, los pasos seguidos en la toma de espectros NIR de para el café es la siguiente:

- Revisar la correcta configuración de los elementos y cableado del equipo.
- Si algún elemento no está correctamente conectado, conectarlo adecuadamente. Por ejemplo, no estar conectada la sonda apropiada.
- Encender el equipo NIR y la lámpara de la sonda. Dejar calentar la lámpara 15-20 minutos.
- Encender el ordenador del equipo NIR y activar el archivo de adquisición de datos de la aplicación del café para la que se van a realizar medidas.
- Tomar la referencia (Spectralon).
- Realizar las medidas NIR sobre las muestras que correspondan para adquirir los espectros NIR correspondientes, limpiando con agua destilada la sonda después de cada muestra y secándola convenientemente, para evitar contaminación entre muestras.
- Realización del método analítico.

Después de tomar los espectros correspondientes a la muestra se ha procedido a la realización del método analítico correspondiente. Lo primero que se hace es tapar el recipiente en el que se deposita el café con el café dentro, para realizar las medidas NIR. De esta forma, se logra aislar al café de la humedad ambiente hasta realizar las medidas del método analítico, aunque se trató de proceder lo más rápidamente posible. Debido al diseño de experimento de cada aplicación de la tecnología NIR, los tiempos de ese diseño experimental de cada aplicación de la tecnología NIR y la alta probabilidad de

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

captura de humedad ambiente por parte del café, cada muestra generada solo se utilizará para una aplicación. En ningún caso se ha utilizado la misma muestra para crear modelos NIR de dos aplicaciones (parámetros).

Relación Arábica/Robusta.

En este caso, se han usado cafés en grano arábica y robusta de distintos orígenes y calidades. En concreto, se han usado café Arábica de Brasil, Nicaragua, Costa Rica y de tres proveedores de Colombia, que es un país muy importante en la exportación de café arábica. Para el caso del café Robusta se ha usado café proveniente de Uganda, Brasil y dos proveedores de Vietnam (calidad alta y baja), Este último país es muy importante en la exportación de café robusta. En total 6 cafés Arábica y 4 Robusta.

Las mezclas se crean mezclando un café arábica con un robusta. En este caso, la suma de la cantidad total del café arábica y el robusta en una muestra debe pesar 9.5 g. aproximadamente. Se han creado mezclas de 9 combinaciones de cafés (4 arábica*3 robusta=12 combinaciones posibles). Para cada combinación entre arábica y robusta se generan alrededor de 8 mezclas, más los cafés arábica y robusta puros. La distribución de las muestras se realiza en función de las muestras ya generadas, tratando de colocarlas en las zonas de composición donde haya poca densidad de muestras. Esto se hace con el ánimo de lograr que los puntos se distribuyan a lo largo de todo el rango de concentraciones arábica/robusta y que las muestras estén equiespaciadas.

Para generar un punto individual de una determinada composición arábica robusta, se pesa en una balanza un recipiente, en el que se contendrán los granos de café verde, y se anota el valor de la pesada. En este caso se ha usado pequeños vasos de plástico. Después se introduce la cantidad de café arábica necesaria para lograr la composición arábica/robusta requerida para el peso total de la muestra, y se pesa en la balanza. Se echa café en el recipiente hasta lograr el valor de peso buscado. Una vez logrado el peso buscado se anota. Seguidamente se añade la cantidad de café robusta, de igual forma que con el café arábica, hasta alcanzar el peso total de la muestra. También, se anota también este valor.

Como se puede ver este método analítico es una excepción respecto al resto de métodos del proyecto. En este caso primero se calcula el método analítico y luego se realizan las medidas NIR. Esto se debe a que la preparación de muestras aporta un valor exacto de la propiedad analítica a predecir (composición Arábica/Robusta), mientras que en los otros casos esto no es así, se determina el valor analítico después de obtener los espectros NIR.

Para calcular el porcentaje de café arábica en las mezclas se han utilizado los pesos apuntados y la siguiente fórmula:

$$\%A = \frac{A - V}{T - V} \times 100$$

Donde *V* es el peso del bote vacío, *A* es el peso del bote con solo el café arábica y *T* es el peso total de la mezcla con los cafés arábica y robusta. En caso de necesitarse el % de café robusta solo hay que aplicar la siguiente fórmula:

$$\%R = 100 - \%A$$

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

A estos desarrollos hay que añadir otra prometedora vía de trabajo que consiste en moler las mezclas creadas de arábica/robusta medirlas también con espectroscopia NIR. La razón de la inclusión de esta vía viene de la lectura de varios artículos, en los que para salvar las dificultades para la espectroscopia que presenta que las muestras sean granos, se muele el café para que las muestras sean más homogéneas. Además, en los análisis realizados a los datos a través de la construcción de modelos matemáticos de predicción de los % de Arábica en mezclas de arábica/robusta de café verde entero se ha observado que los modelos obtenidos tenían poca capacidad predictiva. En este caso, la forma de tomar espectros es similar al café tostado molido. En el recipiente se colocan entre 7 y 8 g de café verde molido. No se dispone de café verde molido por lo que se procede a su molienda. En este caso el tiempo de molienda fueron 2 minutos, divididos en periodos de 20 segundos. Entre cada uno de estos periodos se para el molinillo de café, durante 30 segundos, de café para evitar que se queme el motor del mismo. Mientras el molinillo está parado se desconecta para evitar percances y revuelve el café con una cucharilla dentro del molinillo para evitar que se produzcan zonas con diferente tamaño de partículas y eso genere un tamaño de partículas demasiado heterogéneo. Una vez molido se introduce en un bote de plástico, se cierra el bote y se utiliza parafilm para las juntas de la tapa del bote. Esto es consecuencia de no poder medir los espectros NIR del café verde recién molido porque está caliente por la fricción con las aspas del molino y debe dejarse un tiempo para enfriar. El bote se utiliza para evitar que el café adquiera humedad del ambiente. Una vez transcurridos 20 minutos ya se pueden realizar las medidas NIR. Se introduce la cantidad anteriormente mencionada en el recipiente de medida NIR y se realiza una medida. Se vacía el recipiente, se vuelve a cargar y se realiza otra medida. Este procedimiento se ejecuta 10 veces, de forma que se tienen los esperados 10 espectros, que se promediaran para tener un espectro representativo de la mezcla arábica/robusta.

Humedad de café verde.

Para determinar la humedad del café verde, se utilizarán los mismos tipos de café que en el caso de la aplicación de la determinación de la composición arábica/robusta, los seis tipos arábica y los cuatro de robusta.

En este caso, la clave para lograr tener los datos necesarios es lograr muestras con diferente humedad. Estos siete tipos de café tienen diferente humedad, pero las variaciones entre ellos son pequeñas, ya que todos los cafés verdes tienen una humedad próxima al 10 % en peso. Por esta razón, se han ideado dos formas de modificar la humedad de las muestras de café verde:

- La primera forma sirve para hacer disminuir la humedad. Para ello se ha usado una estufa que estaba a alrededor de 60 ºC. El café es introducido en un plato y en función del tiempo que haya estado dentro de la estufa perderá más o menos agua. Este tiempo puede variar entre un cuarto de hora hasta 4-5 horas. Una vez transcurrido el tiempo dentro de la estufa, se seca el café y se deja enfriar en un desecador. Pasados 20 minutos el café ya está a temperatura ambiente y es puesto dentro de placas Petri y selladas estas con parafilm, para evitar la entrada de humedad exterior. Los granos de café se deben dejar dentro de las placas Petri cerradas porque la perdida de humedad en la estufa es rápida, lo que puede provocar un gradiente de concentración

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

dentro del grano de café, con este periodo de tiempo se trata de homogeneizar, en lo posible, la humedad dentro de los granos de café.

- La segunda forma de variar la humedad ha consistido en utilizar un recipiente cerrado (figura 2.2.3) donde la humedad sea alta. Para ello dentro de este recipiente se mete un pequeño vaso de plástico con algodón humedecido (usar agua líquida puede causar condensaciones indeseadas sobre el café), además de los granos de café en una placa Petri. Después de esto se cierra la tapa del recipiente. Indicar que es importante que el recipiente sea hermético. El café puede estar en el recipiente desde medio día hasta 3-4 días, en función de la humedad en grano de café que se desee, y temperatura ambiente. Una vez terminado el periodo de tiempo se pueden realizar las medidas NIR con este café.

Una vez se han toman los espectros de las muestras, se puede determinar la humedad de las mismas. Para ello como se comentó con anterioridad (medidas NIR para aplicaciones del café), una vez medido los granos de café mediante NIR, se llevan los granos de café en el recipiente de medida NIR a la balanza. Primero se pesa un contenedor de porcelana y se anota su peso. Seguidamente se añaden los granos de café al contenedor y se pesa todo el conjunto, anotándose de nuevo el peso. El recipiente cerámico se introduce en una estufa a 105 ºC durante 24 horas, transcurridas las cuales se saca de la estufa y se coloca a enfriar dentro de un desecador durante 20 minutos. Al finalizar este tiempo se pesa el recipiente con los granos secos de café. Los % de humedad de los granos de café verde se obtienen mediante la siguiente ecuación:

$$\%H = \frac{CVS - R}{CVH - R} \times 100$$

Donde *R* es el peso en gramos del recipiente cerámico, *CVH* es el peso en gramos del recipiente cerámico más los granos húmedos y *CVS* es el peso en gramos del recipiente cerámico más los granos secos.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias INNOVACIÓN

Figura 2.2.3.- Recipiente donde se produce la humidificación de los granos de café.

Humedad de café entero tostado.

En este caso se han usado 14 marcas comerciales de café compradas en supermercado, que serán las mismas que se utilizarán en la determinación de los otros parámetros de café que usen café tostado entero. El procedimiento es similar al caso de la humedad en café verde. Aunque por supuesto existen algunas diferencias que se deben tener en cuenta en el desarrollo de los experimentos:

- En este caso cambia el tipo de café, es café tostado. En un principio se pensó en tostar café verde, pero al final se ha optado por comprar café comercial. La razón fue que existe una gran variedad de cafés en el mercado, con diferentes orígenes y grados de tueste. Además, con un tostado propio difícilmente se iba a lograr café similar a un tostado industrial.
- Las formas de variar la humedad son iguales al café verde, con la salvedad de los tiempos de humidificación y deshumidificación. En este caso, debido a las propiedades físicas de los granos de café tostado, la difusión del agua a través de ellos es más rápida. Así, estos tiempos son alrededor de cinco veces menores que en el caso del café verde.
- El secado en estufa para la determinación de la humedad es igual a 105 ºC, pero en este caso en vez de 24 horas, debe estar en la estufa 16 horas.

Humedad de café tostado molido:

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

En este caso se han usado 15 marcas comerciales de café compradas en supermercado, que serán las mismas que se utilizarán en la determinación de los otros parámetros de café que usen café tostado molido. También, este procedimiento es similar a los dos anteriores con algunas diferencias:

- El café es molido y tostado. Por razones similares al caso anterior se ha optado por marcas comerciales. Además, en este caso si se tuviera que moler café sería difícil lograr el tamaño de grano y la homogeneidad que se logra en el mismo utilizando molinos industriales.
- El café molido es con mucho el que presenta una mayor rapidez en la humidificación y deshumidificación. Este tipo de café es altamente higroscópico. El café molido adquiere humedad ambiente sin necesidad de algodón. En caso de usar algodón húmedo, el café adquiere humedad en cuestión de minutos o unas pocas horas, en función de la temperatura ambiente (cuanta más temperatura más rápido adquiere humedad, ya que el aire tiene una mayor humedad en el habitáculo de humidificación). La retirada de humedad del café en estufa es incluso más rápida, cuestión de minutos.
- El secado en estufa para la determinación de la humedad es a 105 ºC durante 16 horas.
- En este caso, se debe ser especialmente rápido realizando las medidas NIR como durante todo el método analítico debido a que el café molido es muy higroscópico y atrapa la humedad ambiente. Esto como es fácil imaginar es muy perjudicial para el método analítico que es un método gravimétrico.

Color del café tostado entero (Parámetro L y Hue):

En este caso se utilizarán las mismas marcas comerciales que en el caso de la humedad de café tostado entero. Pero existe una salvedad, en este caso solo se puede extraer una muestra por paquete (lote), cuando se utiliza una marca más de una vez debe ser de lotes diferentes. También, una tostadora de café de Asturias cedió café tostado para aumentar la cantidad de muestras del experimento. Esta propiedad es la que se suele usar en la industria cafetera como indicador del grado de tostado. Los parámetros de interés son la luminosidad (L) y el ángulo de Hue en la escale CIE L*a*b*. La luminosidad la aporta directamente el colorímetro, mientras que el ángulo de Hue se calcula a partir de dos parámetros que aporta el colorímetro (a y b) con la siguiente ecuación:

$$Hue = Arc.Tg.\left(\frac{b}{a}\right)$$

Después de tomar los espectros de las muestras, se utiliza un colorímetro para determinar los parámetros de color. El método es similar a la espectroscopia NIR. Primero se toma la referencia del colorímetro, con tres medidas, y otras tres de la corriente de oscuridad (medidas tomadas disparando hacia el aire sin apuntar hacia ningún objeto ni fuente de luz). El colorímetro realiza la calibración con estos dos valores. En este caso, se ha usado el mismo recipiente donde se deposita el café para la obtención de espectros NIR. La forma de proceder consiste en llenar el recipiente de medida NIR con el café, obtener un espectro NIR, y luego con la misma disposición del café realizar una medida con el colorímetro, se apuntan los valores de los parámetros de color, se vacía el recipiente para las medidas NIR. Se vuelve a llenar el recipiente de medida NIR, de forma que se tiene otra disposición de café y se

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

repiten las medidas. Esto se repite 10 veces, de forma que se tienen diez espectros y diez medidas de color, para diez disposiciones de los granos de café en el recipiente de medidas NIR. Cuando se tienen todas las medidas de color realizadas, se calcula la media de las diez medidas para cada uno de los parámetros de color.

Dado que en el transcurso de la construcción de modelos matemáticos de predicción de los parámetros de color de café tostado entero se ha observado que los modelos obtenidos tenían poca capacidad predictiva, sobre todo el parámetro L que depende de la luz visible total que proviene de la muestra. Los parámetros de color se ven muy afectados por la geometría de las muestras, por eso se ha decidido molerlo para homogeneizarlo y hacer uniforme la superficie de las muestras. En este caso, la forma de tomar espectros es similar al café tostado molido. En el recipiente se colocan entre 7 y 8 g de café tostado molido. El tiempo de molienda fueron 2 minutos, divididos en periodos de 20 segundos. Entre cada uno de estos periodos se para el molinillo de café, durante 30 segundos, de café para evitar que se queme el motor del mismo. Mientras el molinillo está parado se desconecta para evitar percances y revuelve el café con una cucharilla dentro del molinillo para evitar que se produzcan zonas con diferente tamaño de partículas y eso genere un tamaño de partículas demasiado heterogéneo. Una vez molido se introduce en un bote de plástico, se cierra el bote y se utiliza parafilm para las juntas de la tapa del bote. Esto es consecuencia de no poder medir los espectros NIR del café tostado recién molido porque está caliente por la fricción con las aspas del molino y debe dejarse un tiempo para enfriar. El bote se utiliza para evitar que el café adquiera humedad del ambiente. Una vez transcurridos 20 minutos ya se pueden realizar las medidas NIR. Se introduce la cantidad anteriormente mencionada en el recipiente de medida NIR y se realiza una medida. En este caso, la forma de actuar difiere del café tostado molido en la determinación de la humedad. Así, se vacía el recipiente, se vuelve a cargar y se realiza otra medida. Este procedimiento se ejecuta 10 veces, de forma que se tienen los esperados 10 espectros, que se promediaran para tener un espectro representativo de la muestra.

Densidad del café tostado entero.

En este caso, se utilizarán las mismas marcas comerciales que en el caso de la humedad de café tostado entero. Pero existe una salvedad, en este caso solo se puede extraer una muestra por paquete (lote), cuando se utiliza una marca más de una vez debe ser de lotes diferentes. También, una tostadora de café de Asturias cedió café tostado para aumentar la cantidad de muestras del experimento. En este caso, a 20 ºC (se debe usar el climatizador para mantener la temperatura), se ha utilizado un picnómetro para medir la densidad de los granos de café tostado (figura 2.2.4). Se ha usado tolueno en vez de glicerina como dicen los primeros artículos sobre el tema porque los granos de café entre el café y la glicerina se forman burbujas que nunca se pueden eliminar por completo (aspecto muy problemático ya que modifica el resultado de la densidad) y hay bibliografía más moderna que lo usa para los mismos fines que este caso. Los pasos a seguir para obtener la densidad del café tostado, son los siguientes:

- Primero se determina la masa del café de la muestra. Se pesa un recipiente vacío y se apunta su peso. Seguidamente se introduce los granos de café tostado en el recipiente y se pesa el

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

conjunto también. La diferencia entre la segunda pesada y la primera pesada es el peso de los granos de café tostado.

- Ahora se pesa el picnómetro con tolueno hasta el enrase.
- Después, se pesa otra vez el picnómetro con el tolueno hasta el enrase, pero con los granos de café tostado dentro.
- Mediante la siguiente ecuación se han obtenido los gramos de tolueno desalojados por la inclusión de los granos de café dentro del picnómetro:

$$Mt = M1 + Mc - M2$$

Donde M1 es la masa en gramos del picnómetro con solo tolueno, Mc es la masa en gramos de los granos de café y M2 es la masa en granos del picnómetro con tolueno y los granos de café dentro.

- El volumen de tolueno desalojado es igual al volumen de los granos de café. Así, mediante la siguiente ecuación se ha calculado la densidad de los granos de café (g/cm^3):

$$Dc = \frac{(Mt \times dt)}{Mc}$$

Donde Mt es la masa en gramos del tolueno desalojado, dt es la densidad del tolueno (0.8623 g/cm^3 a 20 ºC) y Mc es la masa en gramos de los granos de café.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias INNOVACIÓN

Figura 2.2.4.-Picnómetro con café y tolueno pesado.

Pérdida de peso del café tostado entero.

En este caso se utilizarán las mismas marcas comerciales que en el caso de la humedad de café tostado entero. Pero existe una salvedad, en este caso solo se puede extraer una muestra por paquete (lote), cuando se utiliza una marca más de una vez debe ser de lotes diferentes. También, una tostadora de café de Asturias cedió café tostado para aumentar la cantidad de muestras del experimento. Después de realizar los pertinentes espectros NIR de la muestra de café tostado, se empieza el proceso de determinación de la pérdida de peso. Esta determina como el % de cambio para una cantidad predefinida de granos de café crudo respecto al café tostado. En este caso se utilizarán 100 granos de café tostado, para determinar su peso se ha utilizado una balanza. Primero se pesa un recipiente vacío y luego el recipiente con los 100 granos de café. La diferencia entre estos dos valores es el peso de los 100 granos de café. Por las razones dadas en la tarea 2.1, se va a utilizar una referencia para los valores de peso inicial de café verde. Dado que para la determinación de humedad de café verde y el % de Arábica/Robusta se usa café sin tostar, se ha aprovechado para pesar muestras de los cafés verdes (de diferentes orígenes) pesados de la misma forma que en el caso del café tostado (100 granos). En total se

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

pesaron y promediaron 10 cafés de 9 orígenes diferentes (hay dos cafés diferentes de Vietnam y otros tres de Colombia, de diferente proveedor), los mismos que se usaron para la determinación de la humedad del café verde y la relación de % de café arábica en mezclas arábica-robusta. La fórmula usada para obtener el valor de los % de pérdida de peso es:

$$\%PerdidaPeso = \frac{P_{ver} - P_{tos}}{P_{ver}} * 100$$

En esta fórmula P_{ver} es el peso promedio de 100 gramos de las variedades de café verde y P_{tos} es el peso de 100 granos del café tostado analizado.

A modo de resumen, en la siguiente tabla 2.2.2 se pueden ver las muestras generadas y espectros tomados para realizar los modelos de los parámetros para el café.

Tabla 2.2.2.-Muestras usadas para la construcción de modelos matemáticos y espectros para el café.

	Nº Muestras	Nº Espectros
Humedad Verde	127	1270
Humedad Tostado Entero	164	1640
Humedad Tostado Molido	128	1280
Color (L) (Entero)	108	1080
Color (L) (Molido)	57	570
Color (Hue) (Entero)	108	1080
Color (Hue) (Molido)	57	570
Densidad	51	510
Pérdida Peso	51	510
Arábica/Robusta (Entero)	78	780
Arábica/Robusta (Molido)	51	510

ATÚN.

En este caso, se han usado latas y botes de cristal de producto al natural. Se han usado productos de 15 marcas diferentes, con un total de 80 muestras con la siguiente distribución (según etiquetado):

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

- Bonito del norte. 20 muestras.
- Atún claro. 40 muestras. Son el doble de los otros dos grupos, ya que se esperan dos especies distintas bajo este etiquetado.
- Atún. 20 muestras.

El procedimiento consiste en tomar primero los cinco espectros NIR de cada muestra y luego recoger una muestra del producto para su posterior análisis de ADN.

Primero se tomarán las medidas NIR (figura 2.2.5), para ello es necesario acondicionar el lugar de medida colocando un plástico protector para evitar ensuciar la mesa de trabajo. Después se abre la lata o bote y se escurre toda el agua posible dejando seca la superficie de la matriz alimentaria donde se realicen las medidas. Después, los pasos seguidos en la toma de espectros NIR es similar a los casos anteriores:

- Puesta a punto de los equipos NIR para las medidas.
- Medidas de muestras de atunes en cinco lugares de la muestra lo más alejados posible unos de los otros.
- Entre medidas limpiar cuidadosamente la sonda para evitar contaminación de diferentes muestras para el método analítico.
- Toma de muestra para el método analítico.

Figura 2.2.5.- Toma de espectros en muestras de atún.

Una vez tomados los espectros de la muestra se procede con el procedimiento de recogida de muestra de atún. En este caso las muestras serán analizadas por ADN en el departamento de biología funcional de la Universidad de Oviedo. Por indicaciones desde el departamento se ha variado la forma de recoger la muestra inicialmente planeada. Así, en vez de usar el E.Z.N.A. Food DNA kit de Omega Bio-Tek se procederá de diferente forma. Después de medir cada muestra se tomará una muestra de un par de gramos aproximadamente con un bisturí, poniendo cuidado en evitar contaminación cruzada entre muestras. Las hojas del bisturí se reemplazarán en cada toma de muestra. Las muestras se introducirán

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

en tubos de plástico (figura 2.2.6) que se rellenarán con un volumen de etanol al menos cinco veces mayores que el volumen de la muestra. Estos tubos se cerraron y se guardaron en el congelador hasta que fueron llevados para su análisis de ADN para la determinación de la especie.

Figura 2.2.6.- Vial con muestra de atún para posterior análisis de ADN.

Los resultados de los análisis de ADN fueron los siguientes:

- De las 40 muestras de atún claro, 38 fueron identificadas como Yellowfin (*"Thunnus Albacanes"*) y 2 como Melva ("Auxis thazard"). Estas dos muestras no serán incluidas en la creación de modelos. No se encontró ninguna muesta de Bigeye *("Thunnus Obesus"*),
- Las 20 muestras de atún las 20 fueron identificadas como Skipjack ("Katsuwomus Pelanis").
- Las 20 muestras de Bonito del norte fueron identificadas como Bonito del norte (*"Thunnus Albacanes"*).

En la siguiente tabla 2.2.3 se pueden ver las muestras generadas y espectros tomados para realizar los modelos de los parámetros del atún.

Tabla 2.2.3.-Muestras usadas para la construcción de modelos matemáticos y espectros para el atún.

	Nº Muestras	Nº Espectros
Atún	80	400

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

HITO 3.- QUIMIOMETRIA: CREACIÓN DE MODELOS NIR DE PREDICCIÓN E IMPLEMENTACIÓN EN PLATAFORMA SOFTWARE.

En todo modelo quimiométrico de espectroscopia NIR existen una fase de creación del modelo y una fase de uso donde se le saca partido al trabajo realizado. En este hito 3, se ha realizado la fase de creación del modelo quimiométrico. Para esta tarea, se utiliza un conjunto de datos (longitudes de onda vs propiedad física) lo suficientemente extenso y mediante técnicas de optimización se obtienen los parámetros del modelo. Una vez está desarrollado el modelo quimiométrico, ya se puede utilizar para predecir los valores analíticos. Simplemente se presentan los espectros correspondientes a las muestras y el modelo devuelve los correspondientes valores analíticos predichos.

Para la construcción del modelo quimiométrico se han usado los datos generados para tal uso en la tarea 2.2. Esta etapa de optimización de parámetros del modelo consta generalmente de cuatro partes altamente interconectadas, que dado su importancia se procederá a explicar en detalle:

- Pretratamientos espectrales.
- Selección de longitudes de onda.
- Desarrollo de calibraciones.
- Elección del mejor modelo quimiométrico NIR.

TAREA 3.1.- PRETRATAMIENTOS DE DATOS NIR.

El primer paso son tareas comunes en toda creación de modelos matemáticos. Estas comprenden ceración de matrices de datos para poder comenzar la creación de modelos o eliminación de pares de datos donde existe pérdida de parte de la información (espectro o dato analítico). Además, todos los siguientes cálculos se han hecho para los datos en modo reflectancia (datos originales, el espectrómetro adquiere datos de reflectancia difusa) y en modo absorbancia. Para transformar los datos en absorbancia desde reflectancia, se aplica la siguiente ecuación a canal (longitudes de onda) del espectrómetro.

$$Abs = -log(Ref)$$

El primer paso para crear los modelos quimiométricos NIR ha sido aplicar a los datos de los espectros diversas técnicas de pretramiento con la finalidad de acondicionar los espectrogramas para obtener un mejor resultado cuando se le apliquen las técnicas de calibrado. Aunque una clasificación de estos métodos es complicada, dado que muchos de ellos se podrían incluir en varias categorías, se pueden definir las siguientes categorías de métodos utilizados:

- Suavizado: “Movil average”, “Median Filter”, “Savitzky Golay Smoother”, …
- Normalización: “Area Normalization”, MSC (“Multiplicative Signal Correction”), EMSC (“Extended Multiplicative Signal Correction”), SNV (“Standad Normal Variate”), “Autoscale Normalization”, …
- Corrección de línea base y derivación: “Savitzky-Golay”, “Gap-Derivative”, “Norris Gap” y

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

"Detrend", …

Estos métodos han sido aplicados a los datos espectroscópicos, se ha aplicado desde un solo tratamiento hasta tres tratamientos de manera consecutiva. Como resultado se han obtenido los datos de los espectros modificados. La idoneidad de los diferentes pretratamientos viene dada por la capacidad de predicción de los modelos finales.

Para la determinación de modelos quimiométricos se han usado todos los pretratamientos propuestos. Como se verá en la tarea 3.3 (donde se presentan los resultados), la utilización de pretratamientos adecuados de los espectros mejora los resultados de las calibraciones, obteniéndose, generalmente, mejores resultados que cuando no se aplica ninguna clase de pretratamiento.

Detrás de cada modelo quimiométrico realizado, está una larga y exhaustiva búsqueda de la mejor combinación de pretratamientos, hasta tal punto que en duración es lo que más tiempo ocupa en el desarrollo de los modelos matemáticos, por dos razones:

- Si se tiene en cuenta que se pueden aplicar de uno a tres pretratamientos y que hay pretratamientos con hasta tres parámetros que ajustar, las posibles combinaciones son enormes.
- Debe dedicarse una parte muy importante del tiempo de la construcción de modelos matemáticos para espectroscopia NIR (en ocasiones la mayor parte del tiempo) a la elección de la combinación de pretratamientos más adecuada, ya que es vital para que luego los modelos de regresión y clasificación logren resultados satisfactorios. Es numerosa la bibliografía que pone de manifiesto esto, indicando la necesidad de aplicar pretratamientos antes de los más potentes métodos de machine learning.

La idoneidad final de una secuencia de pretratamientos vendrá dada por las técnicas de selección de modelos expuestas en la tarea 3.4, por lo que los resultados de todas las tareas de este hito se expondrán en la tarea 3.4.

TAREA 3.2.- SELECCIÓN DE LONGITUDES DE ONDA.

Después de los pretratamientos se encuentra la etapa de selección de longitudes de onda. Esta etapa de la creación del modelo quimiométrico está justificada porque la información útil relativa a los enlaces químicos implicados en el modelo quimiométrico, generalmente se encuentra localizada en determinadas zonas del espectro.

De los métodos sopesados para su uso en la selección de longitudes de onda en este proyecto, se implementaron en rutinas informáticas y utilizaron dos de ellos para la construcción de modelos quimiométricos a lo largo de la ejecución de esta tarea:

- VIP ("Variable importance in proyection").
- Algoritmos genéticos.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias Innovación

Esta etapa va inmediatamente después de los pretratamientos, y al igual que cualquiera de las tres etapas de la creación de modelos quimiométricos, su idoneidad vendrá dada por la capacidad de predicción del modelo final.

El primero de los métodos de selección de longitudes de onda es el método "Variable Importance in Projection" (VIP), que es un método pensado especialmente para estimar la importancia de las diferentes variables usadas en métodos lineales, por esta razón se usará para la estimación de las longitudes de onda importantes cuando se use el método PLS como modelo de estimación funcional. Este método estima la importancia de cada variable en la proyección usada en el modelo PLS (con todas las variables originales). Con este método, se calcula un parámetro para cada longitud de onda y si este parámetro es mayor que un umbral predefinido (generalmente 1 en bibliografía) se considera que es relevante. Una vez se conocen las longitudes de onda mayores que este umbral, se utilizan para recalcular el modelo quimiométrico. En este proyecto, en vez de estudiar un único umbral se estudiarán diez posibles umbrales entre los valores mínimo y máximo de VIP. El mejor umbral vendrá dado por el error de validación cruzada (RMSECV). Como se explicará en los resultados de este hito en la tarea 3.4, este método no logra superar en ningún caso al método con las variables originales.

El segundo de los métodos de selección de longitudes de onda son los algoritmos genéticos. Estos se pueden usar tanto con métodos lineales como no lineales. Se ha desarrollado íntegramente un algoritmo genético, especialmente pensado espectroscopia NIR, para optimizar las longitudes de onda que aportan información relevante para la construcción de modelos matemáticos para espectroscopia NIR.

El algoritmo está basado en las leyes de la genética donde cada individuo de la población le corresponde un código genético donde están definidas sus características logrará mejores o peores resultados en la realización de una función. En este caso, cada individuo constará de un vector de unos y ceros donde se codifican segmentos de longitudes de onda de estudio (uno/activo significa segmento de longitudes que se usa y cero/inactivo que no se usa).

Este algoritmo consta de las siguientes partes que están relacionadas con las sucesivas fases iterativas que logran obtener mejores resultados con la evolución de la población:

- El algoritmo en una primera parte, importa los datos espectroscópicos y analíticos necesarios de archivos CSV donde se encuentran las matrices de datos.
- Se fijan los parámetros propios del algoritmo genético. Estos son
    - Número máximo de iteraciones (generaciones) del algoritmo genético.
    - Número máximo de iteraciones consecutivas del algoritmo genético sin mejora en la función coste.
    - Número de individuos de la población.
    - Fracción de individuos supervivientes en cada generación. El resto de individuos son evaluados junto a los individuos de la nueva generación para escoger aquellos con la función coste más baja.
    - Tasa de mutación (%). Tantos por ciento de genes que mutan en cada generación.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

  - Número de intervalos en los que se divide el rango de longitudes de onda.
  - Genes activos por individuo en la población inicial. Estos son los genes que presentan valor "1".

- Se produce la inicialización del algoritmo con la generación de la población inicial. La generación de la población inicial consta de una matriz de unos y ceros. Donde las filas son el total de los individuos de la población y las columnas los diferentes segmentos de longitudes de onda en los que se divide el espectro (rango de longitudes de onda). Cada individuo es una posible solución al problema, dicho de otro modo, cada individuo es un posible modelo matemático NIR para el conjunto de longitudes de onda que tiene codificado.
- Esta población inicial se evalúa. Esto consiste descodificar el vector correspondiente de unos y ceros para cada individuo y obtener las longitudes de onda que tiene activas ese individuo. Entonces se crea el modelo que se tiene planificado utilizando solo las longitudes de onda activas en los datos espectroscópicos NIR y los datos analíticos. Estos modelos matemáticos tendrán una función objetivo que hará que un individuo sea mejor que otro en función de su valor.
- En función del valor obtenido de la función coste en cada uno de los individuos en su solución al problema tendrá más o menos posibilidades de aportar sus genes a la siguiente generación de individuos. Se hará un ranking donde los individuos con mejores funciones objetivo tendrán mejor ranking.
- A partir de los datos del ranking se crea la siguiente generación de individuos.
- El primer paso es elegir los progenitores para crear la siguiente generación de individuos. Para ello existen reglas de selección de progenitores en función del valor de la función objetivo de los individuos. Entre las que se probaron en este desarrollo, se encontró que funciona bastante bien la regla en la cual a cada individuo del ranking se le asigna una probabilidad de ser progenitor directamente proporcional a lo alto que se encuentra en el ranking.
- Una vez están seleccionados los individuos se aplican las reglas de cruce entre individuos. En este caso se aplican dos reglas de cruce, con un punto de cruce o con dos puntos de cruce. Con un punto de cruce, quiere decir que los cromosomas (vectores de unos y ceros) de los dos progenitores se cortan en un punto y se intercambian para generar dos nuevos descendientes. En la regla de cruce de dos puntos de cruce se producen dos cortes en los cromosomas y se intercambian la sección central del cromosoma, creándose dos nuevos individuos también.
- El siguiente paso consiste en producir mutaciones en los nuevos individuos de la población (descendientes). El % de mutaciones es un parámetro a fijar de antemano. Una mutación consiste en transformar un uno en un cero, o un cero en un uno. Dicho de otra forma, desactivar o activar un segmento de longitudes de onda.
- Se evalúan los nuevos individuos de la población (descendencia).
- Se ordena la población en función de función objetivo.
- Se eliminan los individuos con peor función objetivo, dejando únicamente el número de individuos de la población inicial, a esto se le llama reemplazo.
- Todo este proceso se repite un número prefijado de veces (generaciones).

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

- El individuo con mejor función objetivo de la última generación aportará la mejor solución y será la mejor combinación de longitudes de onda (mejor modelo matemático NIR).

Dado que este algoritmo conlleva una cantidad importante de subrutinas de código, ha sido necesario probarlo para comprobar su correcto funcionamiento antes de incluirlos en el flujo de creación de modelos quimiométricos. Estas pruebas se realizaron en el año 2019 y se utilizaron los datos grados brix de kiwis generados hasta el momento. Se han utilizado todos los datos en el conjunto de entrenamiento, no se ha utilizado conjunto de test, ya que estos cálculos tienen solo una finalidad para comparar métodos y analizar el testeo del algoritmo genético. Se ha utilizado una validación cruzada 10-fold (10 conjuntos de validación cruzada). El método de estimación funcional será el método PLS. En la tabla 3.2.1 se pueden ver los resultados obtenidos. En esta tabla Nº Iteraciones significa el número de generaciones que tienen lugar en el algoritmo genético, Nº Individuos es el número de individuos que tiene la población, Genes activos es el número de genes que al inicio del algoritmo presentan un valor "1", Fracción supervivencia indica la fracción de individuos que sobreviven a la siguiente generación y Tasa mutación indica la fracción de genes totales que mutan para la siguiente generación.

Como se puede ver en la tabla 3.2.1 y figura 3.2.1, se han probado siete configuraciones diferentes del algoritmo genético, siendo GA_1 una base de parámetros bastante habitual en los ensayos que se puede encontrar en bibliografía, y desde GA_2 hasta GA_7 son variaciones de esta configuración (cambios en negrita en la tabla), donde en cada una de ellas se varía un parámetro del algoritmo genético. Con todas las configuraciones de parámetros se logran bastante mejores resultados que en caso de no usar el algoritmo genético (PLS solo). Así, sin utilizar el algoritmo genético se obtiene un error de 0.86 grados Brix mientras que con las diferentes configuraciones del algoritmo genético se obtienen valores entre 0.68 y 0.71 grados Brix. Otra cualidad interesante y atractiva que se puede ver en este caso de la utilización de algoritmos genéticos es que se obtienen valores similares de error aun variando los parámetros del algoritmo genético dentro de un rango razonable. En vista a estas pruebas y pruebas de testeo de las diferentes partes del algoritmo genético, se puede concluir que funciona correctamente. Por estas razones, cuando se disponga de todos los datos los algoritmos genéticos serán una herramienta útil en la estimación de los modelos predictivos de las diferentes propiedades físicas.

Tabla 3.2.1.- Comparación de los valores de RMSECV para los diferentes modelos obtenidos a partir de siete configuraciones iniciales para el algoritmo genético y la utilización del PLS sin selección de longitudes de onda para los datos sin pretratar de grados Brix de kiwis.

	PLS	**GA_1**	**GA_2**	**GA_3**	**GA_4**	**GA_5**	**GA_6**	**GA_7**
Nº Iteraciones		100	100	**50**	100	100	100	100
Nº Individuos		50	50	50	50	**40**	50	50
Genes Activos		5	5	5	5	5	**10**	5

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias Innovación

Fracción Supervivencia		0.5	0.5	0.5	0.5	0.5	0.5	**0.4**
Tasa Mutación		0.01	0.01	0.01	**0.02**	0.01	0.01	0.01
Nº intervalos		100	**50**	100	100	100	100	100
RMSECV	0.86	0.68	0.71	0.71	0.71	0.69	0.7	0.68

Figura 3.2.1.- Comparación de los valores de RMSECV para los diferentes modelos obtenidos a partir de siete configuraciones iniciales para el algoritmo genético y la utilización del PLS sin selección de longitudes de onda para los datos sin pretratar de grados Brix de kiwis.

La idoneidad final de un conjunto de longitudes de onda vendrá dada por las técnicas de selección de modelos expuestas en la tarea 3.4, por lo que los resultados de todas las tareas de este hito se expondrán en la tarea 3.4.

TAREA 3.3.- DESARROLLO DE MODELOS DE CALIBRACIÓN.

Una vez ejecutados los pasos de pretratamiento y selección de longitudes de onda, llega el turno de las calibraciones propiamente dichas. Lo que se logra mediante una calibración es la relación entre unas variables de entrada o dependientes (valores para las diferentes longitudes de onda) y otras de salida o dependientes (propiedad químico-física). En esta tarea, se usan los datos espectrales, procesados con

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

los pretratamientos y selección de longitudes de onda (en caso de ser necesarios), junto con los datos analíticos asociados para lograr las relaciones funcionales óptimas.

En este proyecto se han usado dos tipos de modelos de predicción, regresión y clasificación (cuantitativo y cualitativo, respectivamente). Con el primer tipo se busca la predicción de un valor real, mientras que con el segundo se busca la predicción del valor de una etiqueta (factor). Estos dos tipos de técnicas se pueden dividir a su vez en métodos lineales y no-lineales. Para este proyecto se han usado una técnica lineal y otro no-lineal. Generalmente, con los métodos lineales se logran resultados satisfactorios, pero en ocasiones debido a la no linealidad de los datos, conviene usar métodos no-lineales que son capaces de abordar estos casos. Estos métodos son computacionalmente más complejos y el tiempo requerido es sustancialmente superior al empleado en los métodos lineales. Así, para los métodos lineales, cuando se fija el número de parámetros y se optimiza el modelo, se presenta una solución unimodal (un único mínimo global), sean cual sean los valores iniciales de búsqueda para los parámetros en la optimización numérica. En el caso de los métodos no-lineales, lo habitual es que existan varios mínimos locales a parte del global, en función de las condiciones iniciales de búsqueda de parámetros óptimos. Otra razón de utilizar también métodos lineales es que, aunque los métodos no lineales tienen una capacidad mayor de estimación de funciones, también es más común obtener sobreajuste de los datos. En resumen, los métodos lineales y los no lineales tienen sus ventajas y desventajas, por eso conviene usar los dos tipos.

Todas estas alternativas han sido usadas en el proyecto para diferentes fines. Las técnicas cualitativas de análisis (regresión) se han usado para predecir todos los parámetros del proyecto con la excepción de la determinación de la especie de atún, donde se han usado técnicas cualitativas (clasificación).

Entre las técnicas de técnicas cuantitativas lineales de análisis multivariante para regresión valoradas para utilizar en este proyecto se encuentraban:

- MLR (“Multiple Linear Regression”).
- PCR (“principal component regression”).
- PLS (“partial least squares”).

Entre estos métodos se ha optado por el método PLS, que se ha implementado en la creación de modelos matemáticos de predicción. La razón principal es la superioridad en los modelos quimiométricos conseguidos con este método respecto a los otros dos métodos, como se puede ver habitualmente en bibliografía. Este es un método lineal básico dentro de la quimiometria. Además, dentro de los métodos lineales es uno de los que obtiene mejores resultados. Consisten en reducir la dimensionalidad de los datos, convirtiendo cientos de variables a unas pocas variables latentes que incluyan la mayor parte de la variabilidad de los datos.

También se han usado e implementado métodos de clasificación en la programación informática de la creación de modelos quimiométricos. Entre las más populares de clasificación lineal de patrones, que se ha valorado para su uso en el proyecto, se encuentran:

- PLS-DA (“Partial Least Squares-Discriminant Analysis”).

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

- SIMCA.

El método elegido ha sido el método PLS-DA ("Partial Least Squares-Discriminant Analysis"). Esta es una variante del método PLS cuando las variables de salida son binarias. Dicho de otro modo, es el método PLS que usa variables "dummy" dependientes. Se ha elegido esta técnica porque después de un análisis de la bibliografía se ha llegado a la conclusión de que es la técnica más usada de todas las técnicas de clasificación lineal en quimiometría.

Respecto a los métodos no-lineales, ha se valorado el uso de dos técnicas no lineales de análisis multivariante para el proyecto, estas técnicas son válidas para regresión y clasificación:

- NN ("Neural Networks"-redes neuronales).
- SVM ("Support Vector Machines"-máquinas de vectores de soporte).

Entre estos dos métodos se ha decidido usar las redes neuronales dado que es hay una mayor variedad de paquetes informáticos por el gran crecimiento que está dándose hoy en día en este tipo de técnicas de estimación funcional. Por lo que se ha implementado en las estrategias de creación de modelos predictivos. Este método es considerado uno de los métodos más importantes (puede que el que más) dentro del "machine learning", así es una de las opciones más usadas cuando se habla de problemas donde es necesaria una estimación funcional. En concreto, dentro de las redes neuronales, la más usada es el perceptrón multicapa, junto con el método de optimización de parámetros conocido como algoritmo de retropropagación. Esta es la opción elegida para este proyecto. Básicamente este tipo de red consiste en capas consecutivas de procesamiento de información. Así, los datos de entrada sufren transformaciones matemáticas hasta obtener los resultados finales. Generalmente, existen tres tipos de capas:

- Capa de entrada: Esta es la capa sensorial. Solo recibe los datos, no procesa información.
- Capas ocultas: Son capas consecutivas, donde la información pasa de una capa hacia la siguiente, procesando la información en cada capa. Normalmente, con una capa oculta suele ser suficiente, ya que una red neuronal con una capa oculta es un estimador de funciones universal.
- Capa de salida: Las salidas de esta capa coinciden son las salidas de la red neuronal (Resultados).

Los modelos matemáticos basados en las redes neuronales implementados aquí, tienen dos etapas:

- Un primer paso que consiste en la reducción de la dimensionalidad de los datos espectroscópicos mediante un algoritmo PLS. De forma que permanecerán las variables latentes obtenidas con este método, logrando de esta forma disminuir la dimensionalidad y aligerar el posterior cálculo con redes neuronales. En un principio se valoró usar PCA en vez de PLS, pero mediante varias pruebas se pudo ver que los resultados con el método PLS eran mejores.
- Un segundo paso donde se usan las variables latentes resultantes del PLS como entradas de la red neuronal. Esta red neuronal es un perceptrón multicapa con una capa oculta. Aquí es donde se estima la relación funcional.

Para las redes neuronales se usará la técnica de Bagging, con la finalidad de atenuar el efecto del

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

sobreajuste de los datos, tan común en las redes neuronales. En la técnica "bagging" se escoge aleatoriamente un número prefijado de datos de entrenamiento con reemplazo para crear cada modelo matemático. El modelo matemático global es el promedio de resultados de todos los modelos matemáticos, de esta forma se evita en gran medida el sobreajuste de datos característico de los modelos individuales. Se ha desarrollado la implementación del emsemble que controla la creación de las redes neuronales.

Para poder testear los modelos matemáticos desarrollados, también se ha desarrollado una herramienta informática para visualizar "in situ" como es el desempeño de un modelo matemático NIR. Se ha incluido este desarrollo inicialmente en esta tarea, en vez de en la tarea 3.4, porque mediante esta prueba de concepto se podrán testear no solo modelos matemáticos optimizados, sino que se podrá testear cualquier modelo de los desarrollados. En concreto, lo que se pretende es desarrollar un programa que cuando el espectrómetro realice los espectros necesarios a la muestra, este programa prediga un parámetro analítico para una aplicación concreta de las desarrolladas y muestre el resultado mediante una ventana en pantalla.

En esta tarea, se ha desarrollado un prototipo donde solo se predice un parámetro para una sola matriz alimentaria. Así, este desarrollo se ha centrado en los grados Brix de los kiwis. En el año 2020, este desarrollo se ampliará (más matrices y más parámetros) y mejorará (interfaz con más opciones y programa más rápido) para el resto de parámetros en el hito 4, ya que serán las interfaces gráficas que se utilizarán en las pruebas de planta piloto.

El funcionamiento interno del programa (figura 3.3.1) es el siguiente. Lo primero que hace el programa es crear un objeto de la clase donde se están pretratamiento, longitudes de onda y parámetros. Después, cuando el espectrómetro genera los espectros necesarios (en el caso del kiwi son dos, uno cada lado de la zona ecuatorial de la fruta), el programa recoge todos estos espectros y genera un espectro promedio representativo de la muestra. Este espectro promedio se pasa como entrada del objeto visto anteriormente, se ejecutan los pretratamientos, se utilizan solo las longitudes de onda necesarias y se aplican los modelos de regresión/clasificación para realizar la predicción del parámetro. Finalmente, se extrae la predicción realizada de dentro del objeto y se muestra en pantalla.

Figura 3.3.1.- Diagrama de funcionamiento interno del programa de visualización de las predicciones de

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

grados Brix.

Para usar el programa de predicción de grados Brix en kiwis se deben seguir los siguientes pasos:

- Tener el espectrómetro NIR funcionando, calibrado con la referencia (Spectralon) y en disposición de realizar medidas.
- Ejecutar el programa de visualización de la predicción de los grados Brix en kiwis.
- Cuando se inicia el programa ya está en disposición de realizar medidas. Así que se toman los dos espectros en lados opuestos de la zona ecuatorial del kiwi.
- Automáticamente se muestra el resultado de la predicción de grados Brix en pantalla para esa muestra.
- Para realizar otra medida hay que presionar "Inicio". Después de se realizan las dos medidas en los lados opuestos del kiwi y se vuelve a mostrar automáticamente el resultado.
- Si se comete un error en la medida por la razón que sea, en cualquier punto del proceso de medidas, solo hay que pulsar el botón "Inicio" para reiniciar las medidas.
- El botón directorio sirve para cambiar el directorio donde se almacenan los espectros utilizados en las predicciones de grados Brix en kiwis.
- Para salir del programa solo se tiene que clicar en la cruz de la esquina superior izquierda de la ventana.

Tarea 3.4.- Elección del mejor modelo quimiométrico.

En esta tarea, se han desarrollado e implementado las herramientas para comparar los diferentes modelos matemáticos construidos (pretratamientos, método de selección de longitudes de onda y método de estimación funcional). La elección del mejor modelo matemático conlleva una serie de pasos y procedimientos que han sido implementados en las rutinas informáticas para crear el modelo matemático para espectroscopia NIR. Estas implementaciones en el software son:

- Zona de elección del método de división de los datos: El procedimiento habitual consiste en dividir los datos en dos conjuntos de datos, entrenamiento y test. La opción elegida para este proyecto es que el conjunto de entrenamiento tenga entre un 80 % de los datos y el de test un 20 %. Con los datos de entrenamiento se construirá el modelo quimiométrico. Si no hay muchos datos se utilizan todos los datos para entrenamiento, esta alternativa se ha usado como finalidad exploratoria de los datos.

  Se han implementado varios métodos de división de datos como dividir los datos conservando el orden original, esto es útil cuando se tienen preparados los conjuntos de entrenamiento y test "ad hoc". División de datos aleatoria, es cuando los datos se distribuyen entre los conjuntos de entrenamiento y test aleatoriamente. Método Kennard-Stone, en este método se tienen en cuenta las distancias entre puntos muestrales para crear el conjunto de entrenamiento. Los análisis durante el proyecto utilizaron el método de Kennard-Stone, pero se observó que este método tiene tendencia a crear conjuntos de test "fáciles de predecir", ya que para el conjunto

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

de test son los datos que no van al conjunto de entrenamiento. Se podría decir que se crea un conjunto muy benévolo. Para subsanar este problema se ha implementado en los algoritmos de cálculo, el método Duplex, que una modificación del método de Kennard-Stone. La gran diferencia es que en este caso es que el conjunto de test se elige a la vez y de igual forma que el conjunto de entrenamiento. De esta forma el conjunto de test tiene datos más difíciles de predecir en general.

- Zona de elección del método de validación cruzada: Para evaluar los resultados de los modelos matemáticos se han usado técnicas de validación cruzada con los datos de entrenamiento. Para ello se divide el conjunto de entrenamiento en "K" particiones y se realizan "K" optimizaciones de parámetros. En cada una de estas optimizaciones se utilizan todas las particiones de datos menos una en la optimización de los parámetros del submodelo, de forma que al final todas las particiones son utilizadas en validación.

  Se han implementado tres tipos de validación. Sin validación cruzada, donde todos los datos se usan en un único entrenamiento y en una única validación. Este método es el que ofrece predicciones del error del modelo perores, pero es el más rápido con mucha diferencia. En este proyecto se ha usado únicamente con finalidad exploratoria. Se suele usar para descartar modelos de forma rápida. Los otros dos tipos de validación si son métodos de validación cruzada. K-fold cross validation, donde se crean "K" particiones iguales de los subconjuntos de entrenamiento, el parámetro "K" se fija de antemano, los valores más habituales son 5 o 10. LOOCV ("Leave-One-Out-Cross-Validation"), aquí se crean tantas particiones como datos de forma que de cada vez solo se usa un dato en la validación.

  En ambos métodos de validación cruzada se realiza la media aritmética de los resultados para obtener un único resultado en la estimación del error del modelo matemático. Para el método k-fold cross validation, también se ha implementado la posibilidad de utilizar varias repeticiones para la validación cruzada. Cuando se generan los K conjuntos usados en la validación cruzada se hace de forma aleatoria, esto puede producir conjuntos donde haya más puntos cerca de determinado valor analítico y menos para otros valores. Si la generación de los k conjuntos se realiza N veces y en cada una de estas N veces se realizan los cálculos correspondientes a la k-fold cross validation, se obtendrán N resultados de la validación cruzada que, si se promedian, se obtendrá un valor más fiable y menos dependiente de la aleatoriedad inherente a realizar una sola validación cruzada. El método LOOCV se ha usado sobre todo con finalidad exploratoria y cuando se tienen pocos datos, mientras el método K-fold se ha usado en la parte final de selección de modelos.

- Implementación de estadísticos para cuantificar los tipos de errores del modelo matemático: Se han incluido dos estadísticos en la rutina informática, RMSECV y RMSEP. RMSECV ("Root Mean Square of Cross Validation"), es el error cuadrático medio que presentan los datos de entrenamiento cuando son utilizados para validar los parámetros del modelo. RMSEP ("Root Mean Square of Prediction"), es el error cuadrático medio que presentan los datos de test que no son utilizados en ningún momento en la optimización del modelo.

  En estos estadísticos se pueden observar diferentes comportamientos del modelo

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

quimiométrico. Por ejemplo, tener un RMSEC bajo quiere decir que el modelo ajusta bien los datos de entrenamiento cuando se usan en la optimización de parámetros. Con esto hay que tener especial cuidado porque si el RMSECV (datos de entrenamiento en validación) es alto quiere decir que tenemos sobreajuste en los datos. La situación óptima es que estos dos valores sean bajos y similares (situación ideal). El RMSEP es una validación totalmente externa del modelo (aquí es donde se utilizan los datos del conjunto de test). Conviene hacer esta validación, porque el estadístico RMSECV del modelo, resultante de la validación cruzada, suele tener aún cierto sesgo. Así, con datos totalmente externos se tiene una idea más aproximada del error de predicción del modelo. Conviene aclarar que para la elección del modelo (diferentes métodos de pretratamiento, selección de longitudes de onda y calibraciones) se debe usar el estadístico RMSECV. Nunca se debe utilizar el estadístico RMSEP en ningún punto de la selección del modelo (número de parámetros, ni clase de modelo), debe ser totalmente externo, para evitar posibles sesgos.

Estos estadísticos no sirven para clasificación de atunes porque están pensados para regresiones. Existen estadísticos apropiados para clasificación entre los que se encuentran precisión, exhaustividad, exactitud, F_1 o ROC AUC. En este proyecto se ha elegido es la exactitud, ya que aporta una idea general de las predicciones correctas. La exactitud es:

$$Exactitud = \frac{Número\ de\ predicciones\ correctas}{Número\ total\ de\ predicciones}$$

También se han realizado los desarrollos de redes neuronales necesarios para lograr tener modelos de clasificación no-lineales.

RESULTADOS DE LOS MODELOS MATEMÁTICOS:

Aquí se presentarán los resultados de este hito 3 (tareas 3.1-3.4). Para ello los modelos matemáticos han sido alimentados con los datos generados en el hito 2.

A modo de resumen, algunos aspectos comunes al trabajo de creación de modelos para todas las propiedades físicas a predecir de todos los modelos de regresión son:

- Existen suficientes datos para crear los correspondientes modelos matemáticos.
- Los conjuntos de entrenamiento y test se han elegido en función del método Duplex. Se ha utilizado un 80% de datos para el conjunto de entrenamiento y un 20% para datos de test. Se ha usado el método Duplex en vez del Kennard-Stone que se venía usando en los dos primeros años del proyecto. La razón se explica en la tarea 3.4.
- Se utilizan todos los pretratamientos mencionados en la tarea 3.1, desde uno solo a tres a la vez.
- Los mejores resultados para cada número de pretatamientos fueron usados para su mejora mediante la selección de longitudes de onda con el método VIP y los algoritmos genéticos (tarea 3.2).
- Para realizar las calibraciones, como método lineal se ha usado el método PLS y como método

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

no lineal las redes neuronales.

- Para el cálculo del error de validación cruzada del método PLS, se ha usado una estrategia k-10. Esto consiste en dividir el conjunto de datos de entrenamiento en 10 conjuntos de forma aleatoria. Para las redes neuronales se ha utilizado el error de los datos "out of the box" del método de bagging que promedia redes neuronales (cada ensemble tiene 50 redes neuronales).

A continuación, se pasa a describir los resultados de cada caso particular. Los resultados se mostrarán siempre de la siguiente forma:

- Primero el efecto de los pretratamientos en los resultados de los modelos matemáticos.
- En segundo lugar, se presentarán los resultados de la selección de longitudes de onda y redes neuronales usando bagging.

Kiwis:

En los kiwis se han realizado modelos matemáticos para 3 propiedades físicas (grados Brix, residuo seco y acidez titulable) a partir de los espectros. En la figura 3.4.1 se presenta los espectros de reflectancia usados para la creación del modelo. Estos espectros son similares a los usados para la creación de modelos matemáticos del residuo seco y la acidez titulable.

Figura 3.4.1.-Espectros de reflectancia usados para la creación del modelo matemático de predicción de grados Brix.

Grados Brix:

En la tabla 3.4.1 y la figura 3.4.2 se presentan los resultados obtenidos para los modelos matemáticos de

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

espectroscopia para el caso de los grados Brix utilizando diferente número de pretratamientos junto con el modelo PLS. En cada una de las tablas, se presentan las mejores combinaciones de pretratamientos para cada número de pretratamientos, tanto para los datos de reflectancia como de absorbancia. En la columna 1 se especifica si los datos son de reflectancia (modo original de adquisición de espectros) o son transformados a absorbancia. La columna 2 de las tabla (Nº Pretrat.) indica el número de pretratamientos consecutivos realizados. Las columnas 3-5 (1 Pretrat., 2 Pretrat. y 3 Pretrat.) están los pretratamientos usados por orden se aplicación a los datos espectroscópicos. En la columna 6 está el número de componentes principales óptimo para el método PLS. Por último, en las tres columnas finales, se encuentran los errores de validación cruzada (RMSECV) y de predicción (RMSEP), y el coeficiente de determinación de validación cruzada (R^2_{CV}).

Tabla 3.4.1.- Resultados de las calibraciones para la estimación de grados Brix mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			17	0.94	0.75	0.93
Reflectancia	1	Gap-Derivative			14	0.95	0.64	0.76
Reflectancia	2	EMSC	Gap-Derivative		15	0.95	0.65	0.80
Reflectancia	3	Detrend	Moving Average	Autoscale Normalization	20	0.95	0.69	0.89
Absorbancia	0	Sin Pretatamiento			17	0.94	0.76	0.96
Absorbancia	1	Gap-Derivative			15	0.95	0.69	0.83
Absorbancia	2	Gap-Derivative	SNV		15	0.95	0.65	0.86
Absorbancia	3	Detrend	Savitzky-Golay (smoothing)	SNV	17	0.94	0.72	0.88

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.2.- Resultados de las calibraciones para la estimación de grados Brix mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

En este caso se puede observar que los resultados de los modelos en los que se aplicaron pretratamientos son mejores que los modelos sin pretratamientos. También se puede ver que los modelos con datos de reflectancia tienen en general mejores errores que los de absorbancia. Siguiendo el criterio del error de validación cruzada, el mejor modelo es con un solo pretratamiento (Gap-Derivative) y datos de reflectancia, con valores de R^2_{CV}=0.95, RMSECV=0.64 y RMSEP=0.76. En la figura 3.4.3 se pueden ver los datos experimentales frente a los datos predichos por este modelo. Se puede observar como el modelo logra predecir con bastante fiabilidad los grados Brix en kiwis.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

Figura 3.4.3.- Grados Brix experimentales frente a grados Brix predichos para el modelo matemático aplicando la mejor secuencia de pretratamientos.

El siguiente paso ha sido, a partir del mejor modelo con pretratamientos, aplicar los métodos VIP y Algoritmos genéticos para la selección de longitudes de onda, para reducir el número de variables iniciales que se aportan el modelo PLS. Los parámetros usados en estos métodos son:

- En el método VIP se distribuyeron 10 umbrales equiespaciados entre los valores VIP mínimo y máximo de las variables. El mejor umbral vendrá dado por el error de validación cruzada (RMSECV).
- El algoritmo genético tiene como parámetros una población de 100 individuos, tasa individuos que sobreviven al cruce de 0.5, tasa de mutación de 0.05, 100 iteraciones y el espectro se divide en 50 intervalos, cada uno de los cuales se codifica como un gen que puede estar activo (1) o inactivo (0).

Por otro lado, también se han aplicado redes neuronales artificiales a la mejor secuencia de pretratamientos obtenida junto al método PLS, en concreto se utilizarán las componentes principales obtenidas con el método PLS. El conjunto de redes neuronales ("bagging") consta de 50 redes neuronales, con 3 neuronas cada una y se utiliza para la construcción de cada red neuronal un 90 % de los datos elegidos con reemplazo. El número de componentes principales del método PLS utilizadas como entradas de la red neuronal es de 17, tres más que el mejor modelo PLS con pretratamientos con el fin de incluir variabilidad importante que no esté incluida en el óptimo de componentes principales

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

del PLS.

En la tabla 3.4.2 y la figura 3.4.4, se presentan los resultados de los métodos anteriormente mencionados además de los resultados de los modelos PLS sin pretratamientos y con la mejor secuencia de pretratamientos. En esta table y figura se muestras los errores RMSECV de validación cruzada (RMSEOOB error "Out-of-bag") y los errores RMSEP del conjunto de test para cada uno de los modelos matemáticos.

Tabla 3.4.2.- Tabla comparativa de las diferentes alternativas de modelos matemáticos de estimación de grados brix.

	RMSECV/RMSEOOB	RMSEP
Sin Pretratamientos	0.75	0.93
Con Pretratamientos	0.64	0.76
VIP	0.64	0.76
Algoritmo Genético	0.58	0.78
Redes Neuronales (RMSEOBB)	0.68	0.87

Figura 3.4.4.- Comparativa de las diferentes alternativas de modelos matemáticos de estimación de grados brix.

Sobre los resultados obtenidos decir que, aunque el algoritmo genético logra el menor valor en el error de validación cruzada, presenta un error de test ligeramente superior al método PLS con pretratamientos. Como se verá más adelante para modelos de otros parámetros y matrices

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

alimentarias, es habitual que el método del algoritmo genético tienda en ocasiones a sobreajustar los datos. Comentar respecto al método VIP que se aplicó y los mejores resultados son los aportados por las variables iniciales, no observándose mejora alguna. Esto es una constante para todos los parámetros y todas las matrices alimentarias como se verá en adelante. En la figura 3.4.5, se presentan los valores de RMSECV y RMSEP para de diferentes umbrales VIP estudiados. Se puede ver que los valores más bajos tanto de validación cruzada como de test corresponden con el valor más bajo de VIP (variables originales). En vista de todos estos resultados, el modelo con pretratamientos y el algoritmo genético son los mejores modelos, aportado resultados similares.

Figura 3.4.5.- Valores de RMSECV y RMSEP para de diferentes umbrales VIP en modelos PLS de grados brix en kiwis.

Residuo Seco:

En la tabla 3.4.3 y la figura 3.4.6 se presentan los resultados obtenidos para los mejores modelos matemáticos de espectroscopia para el caso del residuo seco utilizando diferente número de pretratamientos junto con el modelo PLS.

Tabla 3.4.3.- Resultados de las calibraciones para la estimación de residuo seco (fracción másica) mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			14	0.75	0.76	0.80
Reflectancia	1	Savitzky-Golay			16	0.82	0.65	0.71
Reflectancia	2	Savitzky-Golay	MSC		15	0.83	0.63	0.69
Reflectancia	3	Detrend	EMSC	Savitzky-Golay	18	0.80	0.69	0.72
Absorbancia	0	Sin Pretatamiento			16	0.75	0.77	0.76
Absorbancia	1	Gap-Derivative			16	0.78	0.71	0.77
Absorbancia	2	Gap-Derivative	MSC		16	0.81	0.67	0.73
Absorbancia	3	Detrend	Autoscale Normalization	Savitzky-Golay (smoothing)	19	0.78	0.71	0.73

Figura 3.4.6.- Resultados de las calibraciones para la estimación de grados Brix mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

En el caso de los modelos de reflectancia del residuo seco, los modelos con pretratamientos tienen mejores resultados que sin ellos, mientras que en el caso de la absorbancia se logran mejores resultados

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

en la validación cruzada pero no en los datos de predicción. El mejor modelo tiene dos pretratamientos (Savitzky-Golay Derivative seguido de MSC) y datos de reflectancia, con valores de R^2_{CV}=0.83, RMSECV=0.63 y RMSEP=0.69.

El siguiente paso ha sido, a partir del mejor modelo con pretratamientos, aplicar los métodos VIP y Algoritmos genéticos para la selección de longitudes de onda, para reducir el número de variables iniciales que se aportan el modelo PLS. Los parámetros usados para el caso del residuo seco de los kiwis para estos métodos son:

- VIP. 10 umbrales distribuidos equiespacialmente entre los valores VIP mínimo y máximo de las variables.
- El algoritmo genético tiene como parámetros una población de 100 individuos, tasa individuos que sobreviven al cruce de 0.5, tasa de mutación de 0.05, 100 iteraciones y el espectro se divide en 50 intervalos de igual tamaño.

Por otro lado, también se han aplicado redes neuronales artificiales a la mejor secuencia de pretratamientos obtenida junto al método PLS, en concreto se utilizarán las componentes principales obtenidas con el método PLS. El conjunto de redes neuronales ("bagging") consta de 50 redes neuronales, con 3 neuronas cada una y se utiliza para la construcción de cada red neuronal un 90 % de los datos elegidos con reemplazo. El número de componentes principales del método PLS utilizadas como entradas de la red neuronal es de 18, tres más que el mejor modelo PLS con pretratamientos con el fin de incluir variabilidad importante que no esté incluida en el óptimo de componentes principales del PLS.

En la tabla 3.4.4 y la figura 3.4.7 se presentan los resultados de los métodos anteriormente mencionados además de los resultados de los modelos PLS sin pretratamientos y con la mejor secuencia de pretratamientos para el caso del residuo seco.

Tabla 3.4.4.- Tabla comparativa de las diferentes alternativas de modelos matemáticos de estimación del residuo seco.

	RMSECV/RMSEOOB	RMSEP
Sin Pretratamientos	0.76	0.80
Con Pretratamientos	0.63	0.69
VIP	0.63	0.69
Algoritmo Genético	0.54	0.65
Redes Neuronales (RMSEOBB)	0.64	0.60

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.7.- Comparativa de las diferentes alternativas de modelos matemáticos de estimación del residuo seco.

En este caso, todos los métodos obtienen mejores resultados que los datos originales sin pretratar. Cabe destacar al algoritmo genético que logra el menor error de validación cruzada (RMSECV=0.54), aunque el menor error de test y mejor generalización la logran las redes neuronales (RMSEP=0.6). En la figura 3.4.8, se pueden ver los valores experimentales frente a los predichos para el modelo de redes neuronales.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.8.- Residuo seco experimental frente a residuo seco predicho para el modelo matemático aplicando la mejor secuencia de pretratamientos y utilizando redes neuronales.

Acidez titulable:

En la tabla 3.4.5 y la figura 3.4.9 se presentan los resultados obtenidos para los mejores modelos matemáticos de espectroscopia para el caso de la acidez titulable utilizando diferente número de pretratamientos junto con el modelo PLS.

Tabla 3.4.5.- Resultados de las calibraciones para la estimación de la acidez titulable mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			3	0.26	0.146	0.162
Reflectancia	1	Gap-Derivative			6	0.30	0.141	0.153
Reflectancia	2	Area Normalization	Savitzky-Golay		4	0.29	0.143	0.164

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Reflectancia	3	MSC	Detrend	Moving Average	4	0.28	0.143	0.164
Absorbancia	0	Sin Pretatamiento			4	0.28	0.144	0.165
Absorbancia	1	Detrend			5	0.29	0.143	0.165
Absorbancia	2	Savitzky-Golay	Autoscale Normalization		7	0.28	0.143	0.156
Absorbancia	3	Detrend	Moving Average	EMSC	4	0.27	0.144	0.164

Figura 3.4.9- Resultados de las calibraciones para la estimación de la acidez titulable mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

En este caso, todos los modelos quimiométricos desarrollados presentan resultados pobres. El mejor modelo tiene un pretratamiento (Gap-Derivative) y datos de reflectancia, con valores de R^2_{CV}=0.30, RMSECV=0.141 y RMSEP=0.153. Para la selección de longitudes de onda utilizando VIP y algoritmos genéticos, así como con los modelos de redes neuronales se obtuvieron modelos similares a los obtenidos con pretratamientos seguidos del método PLS (resultados pobres).

Café:

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Para el café, se han realizado modelos matemáticos para 8 propiedades físicas/aspectos (humedad café verde, humedad café tostado entero, humedad café tostado molido, % Arábica, parámetro L del sistema de color CIE L*a*b*, parámetro Hue del sistema de color CIE, densidad del café tostado entero, pérdida de peso del café tostado) a partir de los espectros.

Humedad café verde:

En la figura 3.4.10, se presentan los espectros de reflectancia usados para la creación del modelo quimiométrico que relaciona los espectros con la humedad del café verde. Estos espectros presentan una forma similar a los usados en la creación del modelo matemático para estimar el porcentaje de café arábica en café verde entero, ya que son la misma matriz alimentaria. En la tabla 3.4.6 y la figura 3.4.11 se presentan los resultados obtenidos para los modelos matemáticos de espectroscopia para este caso utilizando diferente número de pretratamientos junto con el modelo PLS.

Figura 3.4.10.-Espectros de reflectancia usados para la creación del modelo matemático de predicción de la humedad del café verde.

Tabla 3.4.6.- Resultados de las calibraciones para la estimación de la humedad de café verde mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

Reflectancia	0	Sin Pretatamiento			14	0.99	3.21E-03	4.31E-03
Reflectancia	1	MSC			14	0.99	2.32E-03	3.60E-03
Reflectancia	2	EMSC	Moving Average		14	0.99	2.34E-03	3.62E-03
Reflectancia	3	MSC	Moving Average	Savitzky-Golay (smoothing)	14	0.99	2.34E-03	3.62E-03
Absorbancia	0	Sin Pretatamiento			13	0.98	3.44E-03	5.34E-03
Absorbancia	1	MSC			13	0.99	2.63E-03	4.14E-03
Absorbancia	2	MSC	Detrend		18	0.99	2.43E-03	3.32E-03
Absorbancia	3	Savitzky-Golay (smoothing)	EMSC	Detrend	18	0.99	2.42E-03	3.33E-03

Figura 3.4.11- Resultados de las calibraciones para la estimación de la humedad del café verde entero mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

Los modelos donde se aplican pretratamientos son mejores que si no se aplicarán los mismos. Esto es tanto para reflectancia como absorbancia, y para los errores de validación cruzada como los de test. Los

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

errores obtenidos para validación cruzada son similares desde 1 a 3 pretratamientos, lo mismo pasa para los errores de test. Los errores obtenidos en reflectancia son similares a los de absorbancia. El error mínimo de validación cruzada se obtiene para los datos de reflectancia con 1 pretratamiento (MSC), con valores de R^2_{CV}=0.99, RMSECV=2.34·10^{-3} y RMSEP=3.62·10^{-3}. En la figura 3.4.12, se pueden ver los datos experimentales de humedad de café verde entero frente a los predichos para este modelo matemático.

Figura 3.4.12.- Humedad de café verde experimental frente a humedad predicha para el modelo matemático aplicando la mejor secuencia de pretratamientos y utilizando PLS.

El siguiente paso ha sido, a partir de los mejores pretratamientos, se han aplicado los métodos VIP, algoritmos genéticos y redes neuronales. Los parámetros usados para estos métodos son:

- VIP. 10 umbrales distribuidos equiespacialmente entre los valores VIP mínimo y máximo de las variables.
- El algoritmo genético tiene como parámetros, una población de 100 individuos, tasa individuos que sobreviven al cruce de 0.5, tasa de mutación de 0.05, 100 iteraciones y el espectro se divide en 50 intervalos de igual tamaño.
- El conjunto de redes neuronales ("bagging") consta de 50 redes neuronales, con 3 neuronas cada una y se utiliza para la construcción de cada red neuronal un 90 % de los datos elegidos con reemplazo. El número de componentes principales del método PLS utilizadas como entradas de la red neuronal es de 17, tres más que el mejor modelo PLS con pretratamientos, con el fin de incluir variabilidad importante que no esté incluida en el óptimo de componentes principales del PLS.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

En la tabla 3.4.7 y la figura 3.4.13 se presentan los resultados de los métodos anteriormente mencionados además de los resultados de los modelos PLS sin pretratamientos y con la mejor secuencia de pretratamientos para el caso de la fracción de humedad en café verde entero.

Tabla 3.4.7.- Tabla comparativa de las diferentes alternativas de modelos matemáticos de estimación de la fracción de humedad en café verde entero.

	RMSECV/RMSEOOB	RMSEP
Sin Pretratamientos	3.21E-03	4.31E-03
Con Pretratamientos	2.32E-03	3.60E-03
VIP	2.32E-03	3.60E-03
Algoritmo Genético	1.67E-03	3.62E-03
Redes Neuronales (RMSEOBB)	2.67E-03	3.47E-03

Figura 3.4.13.- Comparativa de las diferentes alternativas de modelos matemáticos de estimación de la fracción de humedad en café verde entero.

En este caso, todos los métodos obtienen mejores resultados que los datos originales sin pretratar. Cabe destacar al algoritmo genético que logra el menor error de validación cruzada (RMSECV=$1.67 \cdot 10^{-3}$), aunque su error de test es similar a los obtenidos con los pretratamientos únicamente. Las redes neuronales a pesar de tener un error de validación cruzada mayor que otros métodos, presentan el menor error de test y mejor generalización (RMSEP=$3.47 \cdot 10^{-3}$).

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Humedad café tostado entero:

En la figura 3.4.14, se presentan los espectros de reflectancia usados para la creación del modelo quimiométrico que relaciona los espectros con la fracción de humedad del café tostado entero. Estos espectros presentan una forma similar a los usados en la creación de los modelos matemáticos para estimar los parámetros de color (L y Hue), y la densidad y pérdida de peso del café, ya que son la misma matriz alimentaria. En la tabla 3.4.8 y la figura 3.4.15, se presentan los resultados obtenidos para los modelos matemáticos de espectroscopia para este caso utilizando diferente número de pretratamientos junto con el modelo PLS.

Figura 3.4.14.-Espectros de reflectancia usados para la creación del modelo matemático de predicción de la facción de humedad del café tostado entero.

Tabla 3.4.8.- Resultados de las calibraciones para la estimación de la fracción de humedad de café tostado entero mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			12	0.97	3.68E-03	3.14E-03

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Reflectancia	1	EMSC			20	0.98	3.18E-03	3.04E-03
Reflectancia	2	SNV	Savitzky-Golay		12	0.97	3.01E-03	3.07E-03
Reflectancia	3	Savitzky-Golay (smoothing)	EMSC	Detrend	15	0.97	3.15E-03	3.21E-03
Absorbancia	0	Sin Pretatamiento			14	0.97	3.52E-03	3.04E-03
Absorbancia	1	Gap-Derivative			14	0.98	3.26E-03	2.29E-03
Absorbancia	2	SNV	Gap-Derivative		12	0.97	3.03E-03	3.17E-03
Absorbancia	3	Detrend	Moving Average	SNV	18	0.97	3.07E-03	3.13E-03

Figura 3.4.15- Resultados de las calibraciones para la estimación de la fracción de humedad del café tostado entero mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

Los modelos donde se aplican pretratamientos son mejores que si no se aplicarán los mismos para los datos de reflectancia en el caso del error de validación cruzada, para los de absorbancia presentan valores similares. Los errores de test son similares en todos los casos salvo para un pretratamiento y

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

datos de absorbancia que es menor que el resto de casos de datos de absorbancia. El modelo elegido como base de posteriores cálculos vendrá dado por el error de validación cruzada (evitando utilizar los datos de test en la creación de modelos). Así, el error mínimo de validación cruzada se obtiene para los datos de reflectancia con 2 pretratamientos (SNV seguido de Savitzky-Golay), con valores de R^2_{CV}=0.97, RMSECV=$3.01 \cdot 10^{-3}$ y RMSEP=$3.07 \cdot 10^{-3}$.

El siguiente paso ha sido, a partir de los mejores pretratamientos, se han aplicado los métodos VIP, algoritmos genéticos y redes neuronales. Los parámetros usados para estos métodos son:

- VIP. 10 umbrales distribuidos equiespacialmente entre los valores VIP mínimo y máximo de las variables.
- El algoritmo genético tiene como parámetros, una población de 100 individuos, tasa individuos que sobreviven al cruce de 0.5, tasa de mutación de 0.05, 100 iteraciones y el espectro se divide en 50 intervalos de igual tamaño.
- El conjunto de redes neuronales ("bagging") consta de 50 redes neuronales, con 3 neuronas cada una y se utiliza para la construcción de cada red neuronal un 90 % de los datos elegidos con reemplazo. El número de componentes principales del método PLS utilizadas como entradas de la red neuronal es de 15, tres más que el mejor modelo PLS con pretratamientos, con el fin de incluir variabilidad importante que no esté incluida en el óptimo de componentes principales del PLS.

En la tabla 3.4.9 y la figura 3.4.16 se presentan los resultados de los métodos anteriormente mencionados además de los resultados de los modelos PLS sin pretratamientos y con la mejor secuencia de pretratamientos para el caso de la fracción de humedad en café tostado entero.

Tabla 3.4.9.- Tabla comparativa de las diferentes alternativas de modelos matemáticos de estimación de la fracción de humedad en café tostado entero.

	RMSECV/RMSEOOB	RMSEP
Sin Pretratamientos	3.68E-03	3.14E-03
Con Pretratamientos	3.01E-03	3.07E-03
VIP	3.01E-03	3.07E-03
Algoritmo Genético	2.12E-03	2.21E-03
Redes Neuronales (RMSEOBB)	4.05E-03	5.14E-03

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

Figura 3.4.16.- Comparativa de las diferentes alternativas de modelos matemáticos de estimación de la fracción de humedad en café tostado entero.

El algoritmo genético y las redes neuronales presentan resultados antagónicos. Mientras las redes neuronales presentan errores, tanto de validación cruzada como de test, mayores que el resto de modelos, el algoritmo presenta ambos tipos de errores mejor que cualquier otro modelo. Esto es igual tanto para reflectancia como absorbancia, pero el error de reflectancia es ligeramente menor. El modelo de reflectancia del algoritmo genético, presenta valores de RMSECV=$2.12 \cdot 10^{-3}$ y RMSEP=$2.21 \cdot 10^{-3}$. En la figura 3.4.17 se pueden ver los datos experimentales de la fracción de humedad de café tostado entero frente a los predichos para este modelo matemático.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.17.- Humedad de café tostado entero experimental frente a humedad predicha para el modelo matemático aplicando la mejor secuencia de pretratamientos y utilizando PLS.

Humedad café tostado molido:

En la figura 3.4.18, se presentan los espectros de reflectancia usados para la creación del modelo quimiométrico que relaciona los espectros con la fracción de humedad del café tostado molido. En la tabla 3.4.10 y la figura 3.4.19, se presentan los resultados obtenidos para los modelos matemáticos de espectroscopia para este caso utilizando diferente número de pretratamientos junto con el modelo PLS.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.18.-Espectros de reflectancia usados para la creación del modelo matemático de predicción de la facción de humedad del café tostado molido.

Tabla 3.4.10.- Resultados de las calibraciones para la estimación de la fracción de humedad de café tostado molido mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			7	0.99	1.86E-03	1.99E-03
Reflectancia	1	Norris-Gap			10	0.99	1.62E-03	1.71E-03
Reflectancia	2	Detrend	Autoscale Normalization		8	0.99	1.64E-03	1.93E-03
Reflectancia	3	Detrend	Savitzky-Golay (smoothing)	Autoscale Normalization	8	0.99	1.64E-03	1.93E-03
Absorbancia	0	Sin Pretatamiento			9	0.99	1.86E-03	1.92E-03
Absorbancia	1	SNV			8	0.99	1.72E-03	2.03E-03
Absorbancia	2	SNV	Detrend		8	0.99	1.72E-03	2.03E-03

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Absorbancia	3	Savitzky-Golay (smoothing)	Area Normalization	Detrend	10	0.99	1.73E-03	2.28E-03

Figura 3.4.19.- Resultados de las calibraciones para la estimación de la fracción de humedad del café tostado molido mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

Para el caso de la fracción de humedad en café tostado molido, los modelos con pretratamientos y datos de reflectancia presentan mejores resultados que los de absorbancia, tanto para errores de validación cruzada como de test. El mejor modelo matemático es el modelo con 1 pretratamiento (Norris-Gap) y datos de reflectancia que presenta valores similares de validación (RMSECV=$1.62 \cdot 10^{-3}$) y test ($1.71 \cdot 10^{-3}$). En la figura 3.4.20 se pueden ver los datos experimentales de la fracción de humedad de café tostado molido frente a los predichos para este modelo matemático.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.20.- Humedad de café tostado molido experimental frente a humedad predicha para el modelo matemático aplicando la mejor secuencia de pretratamientos y utilizando PLS.

El siguiente paso ha sido, a partir de los mejores pretratamientos, se han aplicado los métodos VIP, algoritmos genéticos y redes neuronales. Los parámetros usados para estos métodos son:

- VIP. 10 umbrales distribuidos equiespacialmente entre los valores VIP mínimo y máximo de las variables.
- El algoritmo genético tiene como parámetros, una población de 100 individuos, tasa individuos que sobreviven al cruce de 0.5, tasa de mutación de 0.05, 100 iteraciones y el espectro se divide en 20 intervalos de igual tamaño.
- El conjunto de redes neuronales ("bagging") consta de 50 redes neuronales, con 3 neuronas cada una y se utiliza para la construcción de cada red neuronal un 90 % de los datos elegidos con reemplazo. El número de componentes principales del método PLS utilizadas como entradas de la red neuronal es de 13, tres más que el mejor modelo PLS con pretratamientos, con el fin de incluir variabilidad importante que no esté incluida en el óptimo de componentes principales del PLS.

En la tabla 3.4.11 y la figura 3.4.21 se presentan los resultados de los métodos anteriormente mencionados además de los resultados de los modelos PLS sin pretratamientos y con la mejor secuencia de pretratamientos para el caso de la fracción de humedad en café tostado molido.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Tabla 3.4.11.- Tabla comparativa de las diferentes alternativas de modelos matemáticos de estimación de la fracción de humedad en café tostado molido.		
	RMSECV/RMSEOOB	RMSEP
Sin Pretratamientos	1.86E-03	1.99E-03
Con Pretratamientos	1.62E-03	1.71E-03
VIP	1.62E-03	1.71E-03
Algoritmo Genético	1.37E-03	1.89E-03
Redes Neuronales (RMSEOBB)	1.68E-03	1.92E-03

Figura 3.4.21.- Comparativa de las diferentes alternativas de modelos matemáticos de estimación de la fracción de humedad en café tostado molido.

En este caso se puede ver como el mejor método desde el punto de vista de la validación cruzada son los algoritmos genéticos, pero tienen uno de los peores errores de test, probablemente por un problema de sobreajuste. El modelo con la mejor secuencia de pretratamientos y método PLS tiene errores bajos y parecidos de validación cruzada (RMSECV=$1.62 \cdot 10^{-3}$) y de test (RMSEP=$1.71 \cdot 10^{-3}$), convirtiéndolo en un modelo atractivo.

<u>% Arábica/Robusta:</u>

En el caso de la estimación de los % de Arábica en mezcla de Arábica/Robusta se han realizado los

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACIÓN PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACIÓN CIENTÍFICA APLICADA Y LA TECNOLOGÍA
iasturias INNOVACIÓN

experimentos con café verde entero y viendo los resultados se ha decidido moler el café con el fin de lograr mejores resultados, de forma que una muestra más homogénea se logre unos espectros más representativos de la muestra.

En la tabla 3.4.12 y la figura 3.4.22, se presentan los resultados obtenidos para los modelos matemáticos de espectroscopia para el caso de la estimación de % de Arábica en mezclas de Arábica/Robusta utilizando diferente número de pretratamientos junto con el modelo PLS. En la tabla 3.4.13 y figura 3.4.23, se presentan los resultados para el caso de la creación de modelos matemáticos de las muestras molidas.

Tabla 3.4.12.- Resultados de las calibraciones para la estimación de los % de café Arábica en mezcla Arábica/Robusta de café verde entero mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			2	0.68	16.44	15.64
Reflectancia	1	Savitzky-Golay			1	0.74	14.59	10.13
Reflectancia	2	EMSC	Savitzky-Golay		1	0.76	14.19	10.69
Reflectancia	3	SNV	Detrend	Savitzky-Golay (smoothing)	2	0.74	14.66	12.08
Absorbancia	0	Sin Pretatamiento			1	0.66	16.93	16.41
Absorbancia	1	Savitzky-Golay			1	0.72	15.29	11.47
Absorbancia	2	EMSC	Savitzky-Golay		1	0.74	14.82	11.92
Absorbancia	3	Autoscale Normalization	Savitzky-Golay (smoothing)	Detrend	3	0.73	14.87	11.59

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

Figura 3.4.22.- Resultados de las calibraciones para la estimación de los % de Arábica en mezclas de Arábica/Robusta de café verde entero mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

Tabla 3.4.13.- Resultados de las calibraciones para la estimación de los % de café Arábica en mezcla Arábica/Robusta de café verde molido mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			9	0.91	9.24	6.32
Reflectancia	1	EMSC			9	0.95	6.75	8.76
Reflectancia	2	Savitzky-Golay	EMSC		7	0.96	6.13	8.97
Reflectancia	3	SNV	Savitzky-Golay (smoothing)	Detrend	10	0.95	6.54	12.12
Absorbancia	0	Sin Pretatamiento			11	0.91	9.21	11.03

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Absorbancia	1	EMSC			12	0.92	8.57	12.36
Absorbancia	2	MSC	Detrend		13	0.94	7.21	13.26
Absorbancia	3	Savitzky-Golay (smoothing)	SNV	Detrend	16	0.95	6.68	12.67

Figura 3.4.23.- Resultados de las calibraciones para la estimación de los % de Arábica en mezclas de Arábica/Robusta de café verde molido mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

Lo primero que se desprende de estos resultados es que la molienda con la consiguiente homogeneización de las muestras logra reducir los errores de los modelos matemáticos de predicción a la mitad. Por eso considera un cambio idóneo y a partir de ahora se continuarán los cálculos con los datos de café molido. Respecto a estos datos decir que los modelos de datos de reflectancia aportan mejores errores que los de absorbancia, especialmente en los errores de test. El modelo escogido entre todos ellos es el modelo con 2 pretratamientos (Savitzky-Golay seguido de EMSC) y datos de reflectancia, que presenta errores de validación RMSECV=6.13 y test RMSEP=8.97. En la figura 3.4.24 se pueden ver los datos experimentales de los % de Arábica en mezcla Arábica/Robusta de café verde molido frente a los predichos para este modelo matemático.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACIÓN PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACIÓN CIENTÍFICA APLICADA Y LA TECNOLOGÍA
iasturias INNOVACIÓN

Figura 3.4.24.- % de Arábica en mezcla Arábica/Robusta de café verde molido experimental frente al valor predicho para el modelo matemático aplicando la mejor secuencia de pretratamientos y utilizando PLS.

El siguiente paso ha sido, a partir de los mejores pretratamientos, se han aplicado los métodos VIP, algoritmos genéticos y redes neuronales a los datos de % de Arábica en mezclas de café verde molido. Los parámetros usados para estos métodos son:

- VIP. 10 umbrales distribuidos equiespacialmente entre los valores VIP mínimo y máximo de las variables.
- El algoritmo genético tiene como parámetros, una población de 100 individuos, tasa individuos que sobreviven al cruce de 0.5, tasa de mutación de 0.05, 100 iteraciones y el espectro se divide en 20 intervalos de igual tamaño.
- El conjunto de redes neuronales ("bagging") consta de 50 redes neuronales, con 3 neuronas cada una y se utiliza para la construcción de cada red neuronal un 90 % de los datos elegidos con reemplazo. El número de componentes principales del método PLS utilizadas como entradas de la red neuronal es de 10, tres más que el mejor modelo PLS con pretratamientos, con el fin de incluir variabilidad importante que no esté incluida en el óptimo de componentes principales del PLS.

En la tabla 3.4.14 y la figura 3.4.25 se presentan los resultados de los métodos anteriormente mencionados además de los resultados de los modelos PLS sin pretratamientos y con la mejor secuencia

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

de pretratamientos para el caso de los % de Arábica en mezclas Arábica/Robusta de café verde molido.

Tabla 3.4.14.- Tabla comparativa de las diferentes alternativas de modelos matemáticos de estimación de los % de Arábica en mezclas Arábica/Robusta de café verde molido.

	RMSECV/RMSEOOB	RMSEP
Sin Pretratamientos	9.24	6.32
Con Pretratamientos	6.13	8.97
VIP	6.13	8.97
Algoritmo Genético	3.47	10.41
Redes Neuronales (RMSEOBB)	6.29	9.47

Figura 3.4.25.- Comparativa de las diferentes alternativas de modelos matemáticos de estimación de los % de Arábica en mezclas Arábica/Robusta de café verde molido.

Respecto a los resultados, decir que, aunque el algoritmo genético presenta un buen error de validación cruzada, el error de test es el más alto. Esto se deba probablemente a un sobreajuste de los datos llevado a cabo por el algoritmo genético. Las redes neuronales aportan resultados similares al modelo PLS con pretratamientos, que es el mejor modelo, con errores de validación cruzada (RMSECV=6.13) y de test (RMSEP=8.97).

Parámetro L (Color CIE L*a*b*):

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

En el caso de la estimación del parámetro L del color se han realizado los experimentos con café tostado entero y viendo los resultados obtenidos se ha decidido moler el café con el fin de lograr mejores resultados, de forma que una muestra más homogénea se logre unos espectros más representativos de la muestra, ya que el parámetro L o luminosidad, depende de la cantidad de luz procedente de la muestra y no de las relaciones entre partes del espectro obtenido. Así, dependiendo de la orientación y posición del grano puede sufrir cambios drásticos (como se comentó en la tarea 2.2).

En la tabla 3.4.15 y la figura 3.4.26, se presentan los resultados obtenidos para los modelos matemáticos de espectroscopia para el caso de la estimación del parámetro L del color para café tostado entero utilizando diferente número de pretratamientos junto con el modelo PLS. En la tabla 3.4.16 y figura 3.4.27, se presentan los resultados para el caso de la creación de modelos matemáticos de las muestras molidas.

Tabla 3.4.15.- Resultados de las calibraciones para la estimación del valor L (espacio de color CIE L*a*b*) en café tostado entero mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			1	0.33	4.42	4.55
Reflectancia	1	Savitzky-Golay (smoothing)			1	0.34	4.38	4.55
Reflectancia	2	MSC	Savitzky-Golay (smoothing)		1	0.34	4.40	4.55
Reflectancia	3	Savitzky-Golay (smoothing)	EMSC	Detrend	1	0.34	4.40	4.55
Absorbancia	0	Sin Pretatamiento			1	0.39	4.21	4.45
Absorbancia	1	Moving Average			1	0.41	4.16	4.45
Absorbancia	2	EMSC	Moving Average		1	0.40	4.19	4.45
Absorbancia	3	Savitzky-Golay (smoothing)	Detrend	EMSC	1	0.40	4.18	4.45

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias INNOVACIÓN

Figura 3.4.26.- Resultados de las calibraciones para la estimación del parámetro L del color del café tostado entero mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

Tabla 3.4.16.- Resultados de las calibraciones para la estimación del valor L (espacio de color CIE L*a*b*) en café tostado molido mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			5	0.99	0.76	0.86
Reflectancia	1	MSC			5	0.99	0.74	1.01
Reflectancia	2	Savitzky-Golay (smoothing)	Area Normalization		5	0.99	0.74	0.99
Reflectancia	3	Detrend	Savitzky-Golay (smoothing)	Area Normalization	5	0.99	0.73	1.03
Absorbancia	0	Sin Pretatamiento			3	0.98	0.95	1.13
Absorbancia	1	Detrend			5	0.98	0.89	1.08

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Absorbancia	2	Savitzky-Golay	Autoscale Normalization		5	0.98	0.79	1.09
Absorbancia	3	Detrend	Autoscale Normalization	Savitzky-Golay (smoothing)	7	0.98	0.83	1.06

Figura 3.4.27.- Resultados de las calibraciones para la estimación del parámetro L del color del café tostado molido mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

Lo primero que se desprende de estos resultados es que la molienda con la consiguiente homogeneización de las muestras logra reducir los errores de los modelos matemáticos de predicción a la quinta parte. Por eso se considera un cambio idóneo y a partir de ahora se continuarán los cálculos con los datos de café tostado molido. En los modelos con pretratamientos de café tostado molido se puede ver que los modelos basados en datos de reflectancia aportan errores más bajos que los de absorbancia, tanto para validación cruzada como de test. En este caso, a pesar de que el mejor error de test lo proporciona el modelo sin pretratamiento, su error de validación cruzada es inferior al modelo PLS de reflectancia con tres pretratamientos. Como la validación cruzada se usa para realizar la elección entre modelos con pretratamientos, este modelo con tres pretratamientos es el elegido para cálculos posteriores. Este modelo presenta errores de validación RMSECV=0.73 y test RMSEP=1.03. En la figura 3.4.28, se pueden ver los datos experimentales del parámetro L del color en café tostado molido frente a los predichos para este modelo matemático.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.28.- Parámetro L del color en café tostado molido experimental frente al valor predicho para el modelo matemático aplicando la mejor secuencia de pretratamientos y utilizando PLS.

El siguiente paso ha sido, a partir de los mejores pretratamientos, se han aplicado los métodos VIP, algoritmos genéticos y redes neuronales a los datos del parámetro L del color de café tostado molido. Los parámetros usados para estos métodos son:

- VIP. 10 umbrales distribuidos equiespacialmente entre los valores VIP mínimo y máximo de las variables.
- El algoritmo genético tiene como parámetros, una población de 100 individuos, tasa individuos que sobreviven al cruce de 0.5, tasa de mutación de 0.05, 100 iteraciones y el espectro se divide en 20 intervalos de igual tamaño.
- El conjunto de redes neuronales ("bagging") consta de 50 redes neuronales, con 3 neuronas cada una y se utiliza para la construcción de cada red neuronal un 90 % de los datos elegidos con reemplazo. El número de componentes principales del método PLS utilizadas como entradas de la red neuronal es de 8, tres más que el mejor modelo PLS con pretratamientos, con el fin de incluir variabilidad importante que no esté incluida en el óptimo de componentes principales del PLS.

En la tabla 3.4.16 y la figura 3.4.28 se presentan los resultados de los métodos anteriormente mencionados además de los resultados de los modelos PLS sin pretratamientos y con la mejor secuencia de pretratamientos para el caso del parámetro L del color en café tostado molido.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Tabla 3.4.16.- Tabla comparativa de las diferentes alternativas de modelos matemáticos de estimación del parámetro L del color en café tostado molido.

	RMSECV/RMSEOOB	RMSEP
Sin Pretratamientos	0.76	0.86
Con Pretratamientos	0.73	1.03
VIP	0.73	1.03
Algoritmo Genético	0.45	1.38
Redes Neuronales (RMSEOBB)	0.97	0.89

Figura 3.4.28.- Comparativa de las diferentes alternativas de modelos matemáticos de estimación del parámetro L del color en café tostado molido.

También en este caso el algoritmo genético presenta el más bajo error de validación cruzada y el más alto de test, probablemente debido a sobreajuste. Las redes neuronales presentan el error más alto de validación cruzada, pero en cambio su error de test es comparable al mejor error de test, el de datos sin pretratamiento. El modelo PLS sin pretratamientos presenta errores bajos y parecidos de validación (RMSECV=0.76) y test (RMSEP=0.86), siendo el modelo más atractivo.

Parámetro Hue (Color CIE L*a*b*):

El caso del parámetro Hue es similar al del parámetro L del color. En este caso también se ha molido el café con ánimo de mejorar los resultados. En la tabla 3.4.17 y la figura 3.4.29, se presentan los resultados obtenidos para los modelos matemáticos de espectroscopia para el caso de la estimación del

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

parámetro Hue del color para café tostado entero utilizando diferente número de pretratamientos junto con el modelo PLS. En la tabla 3.4.18 y figura 3.4.30, se presentan los resultados para el caso de la creación de modelos matemáticos de las muestras molidas.

Tabla 3.4.17.- Resultados de las calibraciones para la estimación del valor Hue (espacio de color CIE L*a*b*) en café tostado entero mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			3	0.60	7.40E-02	7.98E-02
Reflectancia	1	EMSC			2	0.63	7.10E-02	7.20E-02
Reflectancia	2	Savitzky-Golay	MSC		2	0.62	7.23E-02	7.42E-02
Reflectancia	3	Detrend	Savitzky-Golay (smoothing)	MSC	2	0.61	7.27E-02	7.41E-02
Absorbancia	0				3	0.67	6.72E-02	5.71E-02
Absorbancia	1	EMSC			1	0.71	6.27E-02	5.28E-02
Absorbancia	2	MSC	Savitzky-Golay		1	0.68	6.59E-02	5.86E-02
Absorbancia	3	Savitzky-Golay (smoothing)	Detrend	EMSC	1	0.67	6.68E-02	5.92E-02

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.29.- Resultados de las calibraciones para la estimación del parámetro Hue del color del café tostado entero mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

Tabla 3.4.18.- Resultados de las calibraciones para la estimación del valor Hue (espacio de color CIE L*a*b*) en café tostado molido mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			5	0.96	1.82E-02	1.47E-02
Reflectancia	1	Savitzky-Golay			9	0.97	1.41E-02	1.07E-02
Reflectancia	2	Area Normalization	Savitzky-Golay		9	0.98	1.22E-02	1.06E-02
Reflectancia	3	Savitzky-Golay (smoothing)	Detrend	SNV	10	0.98	1.31E-02	1.05E-02
Absorbancia	0	Sin Pretatamiento			6	0.99	1.05E-02	1.20E-02

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Absorbancia	1	Savitzky-Golay (smoothing)			10	0.99	9.43E-03	9.63E-03
Absorbancia	2	Savitzky-Golay (smoothing)	Detrend		8	0.99	9.38E-03	9.41E-03
Absorbancia	3	Savitzky-Golay (smoothing)	Detrend	SNV	5	0.99	1.00E-02	1.23E-02

Figura 3.4.30.- Resultados de las calibraciones para la estimación del parámetro Hue del color del café tostado molido mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

En este caso sucede algo similar al caso anterior, donde después de moler (con la consiguiente homogeneización) el café se mejoran mucho los resultados, aquí también el error se reduce a aproximadamente la quinta parte del obtenido en los modelos de café tostado entero. Por eso en los siguientes cálculos se usará el café molido. Para el café molido se puede ver como la aplicación de pretratamientos mejora los resultados de los modelos de los datos originales, tanto para reflectancia como absorbancia. Los errores de absorbancia son en general mejor que los de reflectancia, especialmente para los datos con 1, 2 pretratamientos o sin pretratamientos. Además, para los casos con 1 y 2 pretratamientos los errores de validación cruzada y test, aparte de ser bajos, son similares. El mejor modelo es el que utiliza 2 pretratamientos y datos de absorbancia, ya que presenta los mejores errores de validación cruzada y test. Este modelo presenta errores de validación RMSECV=$9.38 \cdot 10^{-2}$ y test RMSEP=$9.41 \cdot 10^{-2}$. Este modelo se usará como base de los modelos y cálculos posteriores. En la

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

figura 3.4.31, se pueden ver los datos experimentales del parámetro L del color en café tostado molido frente a los predichos para este modelo matemático.

Figura 3.4.31.- Parámetro Hue del color en café tostado molido experimental frente al valor predicho para el modelo matemático aplicando la mejor secuencia de pretratamientos y utilizando PLS.

El siguiente paso ha sido, a partir de los mejores pretratamientos, se han aplicado los métodos VIP, algoritmos genéticos y redes neuronales a los datos del parámetro Hue del color de café tostado molido. Los parámetros usados para estos métodos son:

- VIP. 10 umbrales distribuidos equiespacialmente entre los valores VIP mínimo y máximo de las variables.
- El algoritmo genético tiene como parámetros, una población de 100 individuos, tasa individuos que sobreviven al cruce de 0.5, tasa de mutación de 0.05, 100 iteraciones y el espectro se divide en 20 intervalos de igual tamaño.
- El conjunto de redes neuronales ("bagging") consta de 50 redes neuronales, con 3 neuronas cada una y se utiliza para la construcción de cada red neuronal un 90 % de los datos elegidos con reemplazo. El número de componentes principales del método PLS utilizadas como entradas de la red neuronal es de 11, tres más que el mejor modelo PLS con pretratamientos, con el fin de incluir variabilidad importante que no esté incluida en el óptimo de componentes principales del PLS.

En la tabla 3.4.19 y la figura 3.4.32 se presentan los resultados de los métodos anteriormente

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

mencionados además de los resultados de los modelos PLS sin pretratamientos y con la mejor secuencia de pretratamientos para el caso del parámetro Hue del color en café tostado molido.

Tabla 3.4.19.- Tabla comparativa de las diferentes alternativas de modelos matemáticos de estimación del parámetro Hue del color en café tostado molido.

	RMSECV/RMSEOOB	RMSEP
Sin Pretratamientos	1.82E-02	1.47E-02
Con Pretratamientos	9.38E-03	9.41E-03
VIP	9.38E-03	9.41E-03
Algoritmo Genético	6.10E-03	1.20E-02
Redes Neuronales (RMSEOBB)	1.72E-02	7.75E-03

Figura 3.4.32.- Comparativa de las diferentes alternativas de modelos matemáticos de estimación del parámetro Hue del color en café tostado molido.

En este caso, también, el algoritmo genético presenta el más bajo error de validación cruzada y el más alto de test, probablemente debido a sobreajuste. Las redes neuronales presentan el error más alto de validación cruzada, pero en cambio su error de test es el mejor. Como se apuntó anteriormente el modelo PLS con la mejor secuencia de pretratamientos presenta errores bajos y similares de test y validación, lo que convierte a este modelo en el modelo más atractivo.

<u>Pérdida de peso (%):</u>

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

En la tabla 3.4.20 y la figura 3.4.33 se presentan los resultados obtenidos para los mejores modelos matemáticos de espectroscopia para el caso de la pérdida de peso del café tostado entero utilizando diferente número de pretratamientos junto con el modelo PLS.

Tabla 3.4.20.- Resultados de las calibraciones para la estimación de la pérdida de peso en café tostado entero mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretrat.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP
Reflectancia	0	Sin Pretatamiento			0.24	1.40	1.50	0.24
Reflectancia	1	Gap-Derivative			0.31	1.32	1.49	0.31
Reflectancia	2	MSC	Savitzky-Golay		0.26	1.37	1.62	0.26
Reflectancia	3	SNV	Detrend	Savitzky-Golay (smoothing)	0.22	1.41	1.62	0.22
Absorbancia	0	Sin Pretatamiento			0.22	1.41	1.52	0.22
Absorbancia	1	Savitzky-Golay (smoothing)			0.28	1.35	1.51	0.28
Absorbancia	2	Savitzky-Golay	EMSC		0.28	1.36	1.61	0.28
Absorbancia	3	Savitzky-Golay (smoothing)	Detrend	MSC	0.30	1.34	1.56	0.30

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

Figura 3.4.33.- Resultados de las calibraciones para la estimación de la pérdida de peso del café tostado entero o mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

La estimación de este parámetro presenta resultados pobres. Todos los modelos presentan errores de validación altos y errores de test más altos aún. Aquí se presentan solo los resultados del método PLS con las mejores secuencias de pretratamientos. Para el caso del resto de métodos los resultados son incluso más pobres, con sobreajustes importantes para el algoritmo genético y las redes neuronales.

Densidad (g/cm^3):

En la tabla 3.4.21 y la figura 3.4.34 se presentan los resultados obtenidos para los mejores modelos matemáticos de espectroscopia para el caso de la densidad (g/cm^3) del café tostado entero utilizando diferente número de pretratamientos junto con el modelo PLS.

Tabla 3.4.21.- Resultados de las calibraciones para la estimación de la densidad en café tostado entero mediante espectroscopia NIR, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivo. Método PLS utilizado en la regresión.

Tipo de Datos	Nº Pretr.	1º Pretrat.	2º Pretrat.	3º Pretrat.	Nº CP	R^2_{CV}	RMSECV	RMSEP

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Reflectancia	0	Sin Pretatamiento			1	0.24	3.00E-02	3.22E-02
Reflectancia	1	Savitzky-Golay			1	0.3	2.88E-02	3.22E-02
Reflectancia	2	Area Normalization	Gap-Derivative		1	0.25	2.98E-02	3.48E-02
Reflectancia	3	SNV	Detrend	Savitzky-Golay (smoothing)	1	0.22	3.04E-02	3.48E-02
Absorbancia	0	Sin Pretatamiento			1	0.21	3.06E-02	3.27E-02
Absorbancia	1	Norris-Gap			1	0.27	2.94E-02	3.26E-02
Absorbancia	2	Savitzky-Golay (smoothing)	Area Normalization		1	0.25	2.98E-02	3.47E-02
Absorbancia	3	Savitzky-Golay (smoothing)	EMSC	Detrend	1	0.25	2.98E-02	3.37E-02

Figura 3.4.34.- Resultados de las calibraciones para la estimación de la densidad (g/cm^3) del café tostado entero o mediante espectroscopia, utilizando datos de reflectancia o absorbancia, de 0 a 3 pretratamientos consecutivos y método PLS.

En este parámetro sucede exactamente lo mismo que en el anterior con errores pobres de validación cruzada y más bajos de test. Aquí se mostrarán los resultados del método PLS con las mejores

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

secuencias de pretratamientos. El resto de métodos aportan resultados menos satisfactorios.

Atún:

Clasificación de especies de atún:

Se han realizado modelos matemáticos para la clasificación de especies de atún a partir de los espectros. En la figura 3.4.35 se presenta los espectros de reflectancia usados para la creación del modelo.

Figura 3.4.35.-Espectros de reflectancia usados para la creación del modelo matemático de clasificación de especies de atún.

En función de estos resultados de la tarea 2.2 se ha decidido crear tres categorías para los modelos de clasificación que corresponden a las especies de atún identificadas en el análisis de ADN:

- Atún claro. Yellowfin (*"Thunnus Albacares"*).
- Atún. Skipjack ("Katsuwomus Pelanis").
- Bonito del norte. (*"Thunnus Alalunga"*).

Las dos muestras de Melva se han retirado de proceso de construcción de modelos, quedando un total de 78 muestras.

El procedimiento empleado consistirá en probar diferentes tipos y secuencias de pretratamientos seguidos de longitudes de onda activas en el método. El mejor modelo encontrado en la búsqueda es el siguiente:

- El método usado es el PLSDA con 10 componente principales.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

- Usa datos de reflectancia.
- Secuencia de dos pretratamientos. Gap-Derivative seguido de SNV.
- Rango de longitudes de onda restringido entre 450-1350 nm.
- Los datos están desbalanceados, dado que los datos de atún claro son el doble que cualquiera de las otras dos clases. Por ello en la construcción del modelo es necesario aplicar una técnica de subsampling que logra igualar artificialmente el número de ejemplos por clase. En este caso se usa "up-sampling" que escoge aleatoriamente (con reemplazo) las clases minoritarias hasta alcanzar el tamaño de la clase mayoritaria.

Se ha usado la misma configuración de datos de reflectancia, secuencia de dos pretratamientos e intervalos de longitudes de onda para crear las redes neuronales (Bagging). Solo que en vez de las 10 componentes principales usadas en el método PLSDA, se usan 13 con la finalidad de captar posible información perdida por el método PLSDA. Los resultados de ambos métodos se muestran en la tabla 3.4.22

Tabla 3.4.22.- Resultados de la clasificación de atunes.

	Validación-Cruzada			**Test**		
	Nº Datos	Nº Aciertos	Exactitud	Nº Datos	Nº Aciertos	Exactitud
PLSDA	62	60	0.984	16	15	0.938
	Validación-Cruzada			**Test**		
	Nº Datos	Nº Aciertos	Exactitud	Nº Datos	Nº Aciertos	Exactitud
Redes Neuronales	62	59	0.968	16	12	0.813

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

Figura 3.4.36.-Exactitud de los modelos de clasificación de atunes.

En la tabla 3.4.22 se pueden ver los resultados de los dos modelos de clasificación (PLSDA y Redes Neuronales). En validación cruzada tienen resultados parecidos, pero el modelo PLSDA tiene errores más bajos de test. Esto lo convierte en el mejor modelo. Sobre los errores cometidos por ambos modelos siempre son iguales, errar en la clasificación entre la clase Albacares (atún claro) y Pelamis (atún). Los datos de la clase Alalunga (Bonito del Norte) se clasifican correctamente.

HITO 4.- VALIDACIÓN DE RESULTADOS.

En este hito se definirán los test de validación de los desarrollos anteriores y se ejecutarán.

TAREA 4.1.- Definición de test de validación en planta piloto.

Lo primero que se ha hecho fue definir las pruebas para testar condiciones de uso y parámetros relacionados con la espectroscopía. Estas son:

Tiempo de integración:

Tiempo que el espectrómetro está capturando fotones. Cuanto más tiempo más fotones captura cada canal del espectrómetro. Si se deja demasiado tiempo abierto se satura el detector, de forma similar al tiempo de apertura del obturador de una cámara fotográfica. Normalmente se usa el blanco para fijar el tiempo de apertura. En el espectrómetro Labspec existe una opción manual para fijar el tiempo de integración y otra automática.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

Número de scans:

Para los diferentes parámetros de las diferentes matrices alimentarias se buscará reducir los tiempos de medida, para ello el primer paso ha sido buscar la relación entre el número de scans realizados por el espectrómetro. Un scan es el número de veces que se realiza el conteo de fotones provenientes de la muestra. Todos los scans realizados se promedian y se obtiene un valor representativo. Se busca reducir el número de scans para reducir los tiempos de medida.

Número de medidas sobre la matriz alimentaria:

Se valorará reducir el número de medidas sobre la matriz alimentaria con la finalidad de reducir tiempos de medida.

Toma de referencias (White):

Se estudiará cuanto tomar referencias. Cuanto tiempo puede pasar entre dos tomas de referencia.

Limpieza:

Se estudiará las formas óptimas de limpieza que logren un balance entre el tiempo de limpieza y la pérdida de tiempo en el proceso.

Después, se implementarán los modelos más prometedores en herramientas informáticas similares a la desarrollada en la tarea 3.3. De esta forma cuando se realicen las medidas se podrá obtener el resultado de los parámetros en pantalla y no se tendrá que esperar a su posterior cálculo en el ordenador. Con esto se logra convertir los modelos matemáticos de predicción en un programa informático que permite el uso directo de estos modelos matemáticos.

Por último, se realizarán pruebas con nuevas matrices alimentarias con la finalidad de testear los modelos con las nuevas condiciones de uso. Para estas pruebas se han analizado los resultados y en función de las capacidades predictivas de los modelos se ha decidido que los siguientes parámetros de las matrices alimentarias tienen capacidad predictiva suficiente para implementar los modelos en los programas de uso y evaluarlas “in situ”. También, se aportan los modelos matemáticos que se implementaran en el programa para cada parámetro dado que se consideraron los más idóneos como modelo final de uso:

- Kiwis:
  - Grados Brix: Datos de reflectancia. PLS con mejor secuencia de pretratamientos obtenido con algoritmo genético.
  - Extracto seco: Datos de reflectancia. Redes neuronales con mejor secuencia de pretratamientos.
- Café:
  - Café verde entero:
    - Humedad: Datos de reflectancia. Redes neuronales con mejor secuencia de pretratamientos.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

- Café verde molido:
  - % Arábica. Datos de reflectancia. PLS con mejor secuencia de pretratamientos.
- Café tostado entero:
  - Humedad: Datos de reflectancia. PLS con mejor secuencia de pretratamientos obtenido con algoritmo genético.
- Café tostado molido:
  - Humedad: Datos de reflectancia. PLS con mejor secuencia de pretratamientos.
  - Parámetro L del color (CIE L*a*b*): Datos de reflectancia. PLS sin pretratamientos.
  - Parámetro Hue del color (CIE L*a*b*): Datos absorbancia. PLS con mejor secuencia de pretratamientos.

- Atún:
  - Clasificación por especies: Datos de reflectancia. PLSDA con mejor secuencia de pretratamientos.

TAREA 4.2.- Validación en planta piloto.

El primer paso en esta tarea ha sido realizar las pruebas de validación de condiciones de uso y parámetros relacionados con la espectroscopía que se realizarán en la planta piloto.

Tiempo de integración:

El espectrómetro Labspec 4 tiene dos modos de adquisición del tiempo de integración óptimo: Manual y automático. El manual consiste en elegir el tiempo de adquisición y pasárselo al programa de control del espectrómetro (Indico Pro). La elección del tiempo de integración se hace de tal forma que la cantidad de luz reflejada de forma difusa por la referencia (Spectralom) sea algo inferior al umbral de conteo de fotones del espectrofotómetro para los canales con mayor número de fotones contabilizados. En este caso se ha podido comprobar que el método automático elige siempre adecuadamente este parámetro por lo que la posibilidad de mejora en este apartado es nula y se continuará funcionando con el método automático.

Número de scans:

Para poder tomar decisiones respecto a este parámetro, primero se debe conocer la relación entre el número de scans y el tiempo requerido para realizarlos. Para ello se cronometró el tiempo para diferentes números de scans y se graficó en la figura 4.2.1.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

Figura 4.2.1.- Número de scans frente al tiempo para el espectrómetro Labspec 4.

Los experimentos iniciales se realizaron para 50 scans por medida del espectrómetro, lo que suponen sobre 6 segundos cada medida. Según se baja el número de scans se disminuye el tiempo, pero aumenta el posible ruido en el espectro. La zona más propensa al aumento de este ruido son los extremos del espectrograma, donde los detectores tienen menos sensibilidad. En este caso se han hecho pruebas y se ha encontrado que hasta 20 scans no se aprecia aumento de ruido alguno para todos los parámetros de todas las matrices estudiadas, por debajo de este umbral se detecta un ligero aumento del ruido en los extremos (sobre todo en el caso del café), por lo que se ha fijado en 20 scans el óptimo. 20 scans son aproximadamente 2.5 segundos, lo que se considera un tiempo adecuado para todos los parámetros y matrices estudiadas.

Número de medidas sobre la matriz alimentaria:

En este caso, se ha valorado cambiar el número de medidas. Para ningún caso se ha encontrado oportuno disminuir el número de medidas, ya que este número fue fijado en el hito 2 para lidiar con la heterogeneidad de las muestras. Existe la posibilidad de aumentar su número, ya que de esta forma su promedio aportaría un espectro representativo de las muestras, pero aumentaría el tiempo de toma de espectros. Por estas razones se ha mantenido el número de espectros realizados por muestra que se decidió en el hito 2.

Toma de referencias (White):

Se ha analizado la necesidad de temporal de toma de referencias (Spectralom) y se ha encontrado un

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

patrón de comportamiento de la luz de la sonda del espectrómetro. En función de lo observado este patrón es siempre el mismo. Según el fabricante, antes de realizar un blanco para empezar a medir se debe esperar a que la luz se estabilice, y para esto son necesarios 15 minutos. Según nuestra experiencia 20 minutos es mejor que 15 minutos. Pero aun así la luz de la lámpara sigue evolucionando y conviene realizar dos blancos más en la primera hora de funcionamiento, uno cada media hora. Esto se debe a que la intensidad lumínica durante esta primera hora sigue sufriendo cambios, que se pueden observar en los espectrogramas como desplazamientos (gaps) entre los detectores. En la figura 4.2.2 se puede ver este efecto en un espectro de kiwi. Los saltos se pueden ver, a 1000 y 1800 nm, dentro de círculos rojos. Así en el espectrograma se observa un salto entre las regiones de dos detectores. Después de esta hora inicial, la luz se estabiliza y solo es necesario tomar la referencia cada hora.

Figura 4.2.2.-Espectro donde se pueden ver los saltos debidos al cambio de detectores en el espectrómetro.

Limpieza:

La limpieza necesaria depende mucho de la matriz alimentaria estudiada. Para cada una de ellas se ha decidido que las siguientes directrices marcan un mínimo de limpieza de la sonda:

- Kiwis: Solo será necesario pasar una gasa por la superficie de la sonda del espectrómetro. En caso de estar el kiwi húmedo, se debe asegurarse que se retira adecuadamente la humedad.
- Café Verde entero: Solo será necesario pasar una gasa por la superficie de la sonda del espectrómetro, en caso de detectar suciedad. Aunque esta suciedad se da rara vez.
- Café Verde molido: Solo será necesario pasar una gasa por la superficie de la sonda del espectrómetro para retirar pequeñas partículas de café verde.
- Café tostado entero: Solo será necesario pasar una gasa por la superficie de la sonda del

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

espectrómetro, en caso de detectar suciedad. Aunque esta suciedad se da rara vez.

- Café Tostado molido: Solo será necesario pasar una gasa por la superficie de la sonda del espectrómetro para retirar pequeñas partículas de café tostado molido entre medidas de la misma muestra. La sonda, se debe limpiar con agua entre muestras diferentes y secar con una gasa.
- Atún: Retirar con una gasa suciedad de la superficie entre medidas de la misma muestra. La sonda, se debe limpiar con agua entre muestras diferentes y secar con una gasa.

El siguiente paso ha sido implementar los modelos más prometedores, fijados en la tarea 4.1, en los visores que traducen espectros en predicciones analíticas (figura 4.2.3). Para ello se usa la herramienta informática desarrollada en la tarea 3.3. El procedimiento consiste en cambiar la calibración de grados Brix en kiwi por la que corresponda. El programa hace pasar los espectros obtenidos por esos cálculos y se obtiene la predicción del parámetro buscado en pantalla después de escanear la muestra con el espectrómetro. Se ha creado un programa genérico que permite la carga de diferentes parámetros para las diferentes matrices alimentarias, siempre de uno en uno. Existe la posibilidad de modificar el programa para que muestre más de un parámetro a la vez por matriz alimentaria (como podría ser el caso de los kiwis) pero se ha decidido que en este caso para fines de test y demostrativos es suficiente mostrarlos de uno en uno.

Figura 4.2.3.- Ejemplo de aplicación de predicción “in situ” (grados Brix) de parámetro a partir de

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias Innovación

medidas espectrales.

El último paso ha sido utilizar matrices alimentarias para predecir los parámetros especificados en la tarea 4.1 con las condiciones fijadas en esta misma tarea. Para ello se analizarán entre 15 y 20 muestras nuevas para cada parámetro. En la tabla 4.2.1 se muestran los parámetros estudiados, el número de muestras utilizadas para cada parámetro y el estadístico RMSEP para la predicción de cada parámetro. Los métodos analíticos son exactamente los mismos que se expusieron y utilizaron en el hito 2.

Tabla 4.2.1.- Resultados de los test finales de validación para los diferentes parámetros.

Kiwis		
Parámetro de estudio	Número de muestras	RMSEP
Grados Brix	20	0.69
Extracto Seco	20	0.68
Café verde entero		
Parámetro de estudio	Número de muestras	RMSEP
Humedad	15	3.20E-03
Café verde molido		
Parámetro de estudio	Número de muestras	RMSEP
% Arábica	15	9.17
Café tostado entero		
Parámetro de estudio	Número de muestras	RMSEP
Humedad	20	2.20E-03
Café tostado molido		
Parámetro de estudio	Número de muestras	RMSEP
Humedad	20	1.75E-03
Color (parámetro L)	20	0.79
Color (parámetro Hue)	20	1.04E-02
Atún		
Parámetro de estudio	Número de muestras	Exactitud_Test
Clasificación especies	15	0.93

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

Todas las muestras se generaron de igual forma que en el hito 2. Tratando de abarcar la suficiente variabilidad que en el hito 2 pero a menor escala de datos. Con el café, para la humedad se generaron diferentes humedades artificialmente de igual forma que anteriormente. Para el resto de propiedades estudiadas se adquirieron muestras nuevas para los test. En los kiwis, es la única matriz donde no se pudo reproducir parte de la variabilidad, ya que el kiwi es estacional y no se pudo conseguir kiwi inmaduro. Los ensayos se hicieron con 20 kiwis maduros. Para el caso del atún se han adquirido 5 muestras de atún claro, 5 de atún y otras 5 de bonito del norte.

El número de sesiones de medidas es en función del número de muestras, solo se midieron 5 muestra por sesión. Así, si se midieron 15 muestras se necesitaron 3 sesiones y con 20 muestras se necesitaron 4 sesiones.

Los resultados como se puede observar son satisfactorios. Como curiosidad se pude observar que el error de test es en general algo inferior al error de test obtenido en el hito 3. Esto se puede deber a que en este caso los datos son todos nuevos, no se escogieron para el conjunto de test utilizando el algoritmo Duplex que suele escoger datos para el conjunto de test más extremos y en consecuencia suelen ser más difíciles de predecir.

4. ACTIVIDADES SUBCONTRATADAS

4.1 Especifique las actividades concretas realizadas por cada una de las entidades subcontratadas (si las hubiere), indicando si se corresponden con las previstas en la solicitud o si se han producido desviaciones.
4.2 Valore el grado de cumplimiento de las entidades subcontratadas en la ejecución de sus actividades.

No procede.

5. HITOS ALCANZADOS. GRADO DE CUMPLIMIENTO

5.1 Indique los diferentes hitos alcanzados, relacionándolos con la planificación inicial del proyecto. Detalle los indicadores de seguimiento y entregables obtenidos en el periodo justificado.
5.2 Declare el % estimado de cumplimiento en cada uno de ellos.

HITO 1.- Especificaciones Técnicas de Usuario.

Grado de cumplimiento respecto al total del proyecto: 100 %.

Indicadores de seguimiento: Establecer límites para las propiedades estudiadas del café, kiwi y diferencias entre especies de atún.

Indicadores de seguimiento: Establece condiciones de producción en las industrias alimentarias involucradas.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología
iasturias Innovación

HITO 2.- Diseño de experimento y adquisición de datos.

Grado de cumplimiento respecto al total del proyecto: 100 %.

Indicadores de seguimiento: Establecer diseño experimental para la creación de las calibraciones.

Indicadores de seguimiento: Aplicación efectiva del diseño experimental.

HITO 3.- Quimiometría: Creación de modelos NIR de predicción e implementación en plataforma software.

Grado de cumplimiento respecto al total del proyecto: 100 %.

Indicadores de seguimiento: Utilización de por lo menos cinco tipos de pretratamientos en la creación de la aplicación quimiométrica.

Indicadores de seguimiento: Utilización de por lo menos dos métodos de selección de longitudes de onda en la creación de la aplicación quimimétrica.

Indicadores de seguimiento: Utilización de por lo menos un método lineal y otro no lineal en la creación de la aplicación quimiométrica.

Indicadores de seguimiento: Implementación y uso de los métodos de selección de modelos.

HITO 4.- Validación de resultados.

Grado de cumplimiento respecto al total del proyecto: 100 %.

Indicadores de seguimiento: Establecer plan de tests para validación.

Indicadores de seguimiento: Ejecución del plan de tests para validación.

6. RESULTADOS CONSEGUIDOS

Enumerar brevemente, pero con claridad y precisión, los resultados obtenidos, haciendo referencia expresa a los resultados inicialmente previstos.

Hito 1. Especificaciones Técnicas de Usuario.

Tarea 1.1.- Especificaciones de las matrices alimentarias.

Se han estudiado las diferentes matrices alimentarias, centrándose en los diferentes aspectos que guardan relación con los parámetros estudiados. Se han buscado rangos de valores para estos parámetros para los casos donde proceda.

Tarea 1.2.- Especificaciones de usuario y operacionales.

Se han estudiado los distintos procesos de producción y se conocen los posibles problemas y posibilidades que conllevan en el diseño de experimentos y la implementación de la tecnología NIR.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

HITO 2.- Diseño de experimento y adquisición de datos.

Tarea 2.1.- Diseño del protocolo experimental para asegurar la exactitud y la confiabilidad.

Se ha realizado una profunda revisión bibliográfica para la aplicación NIR a las aplicaciones buscadas. Se han fijado las fuentes de variabilidad que afectan a los diferentes parámetros buscados y sobre los que se genera el diseño experimental. También, se han estudiado aspectos de espectroscopia NIR que influirán en el experimento.

Tarea 2.2.- Implementación de diseño experimental de la adquisición de espectros.

Se han generado las muestras necesarias conforme el diseño experimental. Cada muestra generada tiene asociada varios espectros (su número es función del parámetro de estudio) y los datos analíticos a predecir.

HITO 3.- Quimiometría: Creación de modelos NIR de predicción e implementación en plataforma software.

Tarea 3.1.- Pretratamientos de datos NIR.

Se han implementado y utilizado los pretratamientos de datos de espectroscopia NIR.

Tarea 3.2.- Selección de longitudes de onda.

Se han implementado y utilizado los métodos de selección de longitudes de onda de datos de espectroscopia NIR.

Tarea 3.3.- Desarrollo de modelos de calibración.

Se han implementado y utilizado las técnicas de análisis funcional lineales y no lineales para la predicción de los parámetros de interés de este proyecto.

Tarea 3.4.- Elección del mejor modelo quimiométrico.

Se han implementado y usado todos los algoritmos y métodos matemáticos necesarios para la selección de modelos matemáticos de espectroscopia NIR.

Se han ejecutado y obtenido los modelos quimiométricos completos (resultado de todos los pasos de este hito) para todos los parámetros de todas las matrices de interés.

HITO 4.- Validación de resultados.

Tarea 4.1.- Definición de tests de validación en planta piloto.

Se han valorado y definido todos los aspectos de interés en la puesta a punto de la implementación de los modelos de predicción. Se han establecido los test de validación de los modelos matemáticos.

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Gobierno del Principado de Asturias
FICYT Fundacion para el Fomento en Asturias de la Investigacion Cientifica Aplicada y la Tecnologia
iasturias Innovación

Tarea 4.2.- Validación en planta piloto.

Se han estudiado todos los aspectos de interés en la puesta a punto de la implementación de los modelos de predicción. Se han ejecutado los test de validación de los modelos matemáticos.

7. RESULTADOS PREVISTOS NO CONSEGUIDOS

Indique y explique las causas.

No procede.

8. INFORMACIÓN ADICIONAL

Detalle aquí cualquier otra información que desee hacer constar y que considere relevante para informar de la ejecución del proyecto.

9. CUMPLIMIENTO DE LAS OBLIGACIONES DE DIFUSIÓN Y PUBLICIDAD

En virtud del Apdo. 9 de la convocatoria:

1- Se ha instalado una placa conmemorativa en la que se informa de la ayuda recibida	☒	SI	☐	NO		
2- Se ha informado a proveedores y terceros con los que se ha tenido relación con motivo del desarrollo del proyecto/estudio de la financiación del Fondo Europeo de Desarrollo Regional. En particular se ha recogido la referencia a la financiación europea:						
• En el sitio o portal web del beneficiario, en caso de que disponga de él, identificando claramente el proyecto/estudio subvencionado	☒	SI	☐	NO	☐	NO DISPONE
• En la formalización del encargo a los proveedores externos de I+D	☐	SI	☐	NO	☒	NO PROCEDE
• En los acuerdos de colaboración entre empresas	☐	SI	☐	NO	☒	NO PROCEDE
• En los Estudios de Viabilidad Técnica y en los Proyectos de Investigación Industrial y Desarrollo Experimental subcontratados a proveedores externos de I+D, independientemente del soporte material y/o virtual de los mismos	☐	SI	☐	NO	☒	NO PROCEDE
• En cualesquiera otros documentos elaborados con motivo del proyecto (contratos de subcontratación, informes de auditores, etc.)	☐	SI	☐	NO	☒	NO PROCEDE

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN

• En una comunicación específica, de la que debe quedar constancia de su recepción, a los miembros del personal propio, cuyos costes, hayan sido considerados gastos subvencionables del proyecto	☒	SI	☐	NO	NO PROCEDE	
• En el caso de aparatos, equipos, prototipos, etc., se han fijado a los mismos, etiquetas adhesivas con la referencia a la cofinanciación europea	☐	SI	☐	NO	☒	NO PROCEDE

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
GOBIERNO DEL PRINCIPADO DE ASTURIAS
FICYT FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA
iasturias INNOVACIÓN