



Universidad Pública de Navarra  
Nafarroako Unibertsitate Publikoa

**ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA Y  
BIOCIENCIAS**

**NEKAZARITZAKO INGENIARITZAKO ETA BIOZIENTZIETAKO GOI  
MAILAKO ESKOLA TEKNIKOA**

*APLICACIÓN DE IMÁGENES HIPERESPECTRALES (HSI-NIR) PARA LA CLASIFICACIÓN DE HOJAS DE  
PATATA DE ACUERDO A LOS EFECTOS DE LA FERTILIZACIÓN NITROGENADA*

presentado por

*HELENA ZOZAYA VELA (e)k*

*aurkeztua*

MÁSTER UNIVERSITARIO EN INGENIERÍA AGRONÓMICA

*NEKAZARI INGENIERITZA UNIBERTSITATE MASTERRA*

Marzo, 2022 / 2022, Martxoa



El Trabajo Fin de Estudios titulado:

Grado amaierako lan hau:

**APLICACIÓN DE IMÁGENES HIPERESPECTRALES (HSI-NIR) PARA LA CLASIFICACIÓN DE HOJAS DE PATATA DE ACUERDO CON LOS EFECTOS DE LA FERTILIZACIÓN NITROGENADA**

Presentado por:

Nork aurkeztua:

*Helena Zozaya Vela*

Para optar al Grado en:

Grado hau eskuratzeko:

*Máster Universitario en Ingeniería Agronómica*

Ha sido realizado en la Escuela Técnica Superior de Ingeniería Agronómica de la Universidad Pública de Navarra, en el Departamento de:

Nafarroako Unibertsitate Publikoko Nekazaritza Ingeniarien Goi Mailako Eskola Teknikoko sail honetan:

*Departamento de Ingeniería*

Bajo la dirección del Dr./Dra.

Noren zuzendaritzapean:

*Ainara López Maestresalas y Silvia Arazuri Garín*

En Pamplona, el día:

Iruñean, egun honetan:

*13 de febrero de 2022*

ZOZAYA  
VELA  
HELENA -  
73449237W  
Firmado digitalmente por ZOZAYA VELA HELENA - 73449237W Fecha: 2022.02.14 09:43:41 +01'00'

Fdo./Stua.:

**Helena Zozaya Vela**  
Estudiante / ikaslea

LOPEZ  
MAESTRESALAS  
AINARA - 44622966E  
Firmado digitalmente por LOPEZ MAESTRESALAS AINARA - 44622966E Fecha: 2022.02.13 17:38:47 +01'00'

Fdo./Stua.:

**Ainara López Maestresalas/Silvia Arazuri Garín**  
Director/a / zuzendaria

ARAZURI  
GARIN SILVIA  
- DNI  
44615500P  
Firmado digitalmente por ARAZURI GARIN SILVIA - DNI 44615500P Fecha: 2022.02.13 20:01:48 +01'00'



## **RESUMEN**

En los últimos años, los niveles de CO<sub>2</sub> en la atmósfera se han incrementado exponencialmente provocando un aumento de los estreses abióticos sobre los cultivos. Así, resulta fundamental emplear nuevas tecnologías que permitan identificar qué cultivares responden mejor frente a las consecuencias del cambio climático. Una de las técnicas más utilizadas en la actualidad son las imágenes hiperespectrales (HSI). Teniendo en cuenta la gran importancia mundial del cultivo de patata, en este estudio se han empleado HSI en el rango infrarrojo cercano con el objetivo de evaluar su potencial para identificar la respuesta, en hojas de patata, a la deficiencia de nitrógeno. Esto permitiría, desde el punto de vista económico, medioambiental y social, realizar una fertilización eficiente, acelerar el progreso en la mejora de cultivos y contribuir a la seguridad alimentaria.

Se tomaron imágenes de 240 hojas de plantas de patata (6 variedades diferentes) sometidas a 2 clases de tratamientos de fertilización distintos. Mediante un análisis discriminante de mínimos cuadrados parciales (PLS-DA) se desarrolló un modelo capaz de clasificar correctamente en su clase el 53,06% de las muestras. Por su baja precisión, se concluyó que el empleo de HSI no fue efectivo para diferenciar las hojas sometidas a un déficit de nitrógeno de las cultivadas en condiciones óptimas. Tras contrastar los resultados con el método de referencia, se observó que todas las plantas respondieron igual, independientemente del tratamiento al que estuvieran sometidas. Así, realmente, las hojas de ambas clases no se diferenciaban químicamente y, por ende, se obtuvieron dichos resultados.

## **PALABRAS CLAVE**

Imágenes hiperespectrales, análisis PLS-DA, cambio climático, nitrógeno, patatas

## **ABSTRACT**

In recent years, CO<sub>2</sub> levels in the atmosphere have increased exponentially, causing an increase in abiotic stresses on crops. Thus, it is essential to use emerging technologies to identify which cultivars respond better to the consequences of climate change. One of the most widely used techniques today is hyperspectral imaging (HSI). Taking into account the great global importance of potato cultivation, HSI in the near infrared range has been used in this study to evaluate its potential to identify the response of potato leaves to nitrogen deficiency. This would allow, from an economic, environmental and social point of view, an efficient fertilization, accelerating progress in breeding programs and contributing to food security.

For this purpose, images were taken of 240 leaves of potato plants (6 different varieties) subjected to 2 different fertilization treatments. A partial least squares discriminant analysis (PLS-DA) was used to develop a model capable of correctly classifying 53.06% of the samples. Due to its low accuracy, it was concluded that the use of HSI was not effective in differentiating leaves subjected to nitrogen deficit from those grown under optimal conditions. After comparing the results with the reference method, it was observed that all plants responded the same, regardless of the treatment to which they were subjected. Thus, in fact, the leaves of the two classes did not differ chemically and, therefore, these results were obtained.

## **KEY WORDS**

hyperspectral imaging, PLS-DA analysis, climate change, nitrogen, potatoes

# ÍNDICE

1.	INTRODUCCIÓN Y OBJETIVO .....	1
2.	REVISIÓN BIBLIOGRÁFICA .....	2
2.1.	Cambio Global.....	2
2.1.1.	Causas y consecuencias del cambio climático .....	2
2.1.2.	Repercusión del Covid-19 en el cambio climático .....	5
2.2.	Interrelación del cambio climático y la agricultura.....	6
2.2.1.	Efectos de la agricultura en el cambio climático .....	6
2.2.2.	Efectos del cambio climático en la agricultura .....	7
2.3.	Imágenes Hiperespectrales (HSI) .....	10
2.3.1.	Base teórica.....	11
2.3.2.	Elementos del sistema de imágenes hiperespectrales .....	12
2.3.3.	Métodos de detección y adquisición de imágenes hiperespectrales .....	13
2.3.4.	Etapas del proceso de análisis .....	15
2.3.5.	Ventajas y desventajas.....	18
2.4.	Quimiometría .....	19
2.4.1.	Análisis cualitativo.....	19
2.4.2.	Análisis cuantitativo .....	20
2.5.	Aplicaciones HSI en el cultivo de patata. ....	21
3.	MATERIALES Y MÉTODOS .....	22
3.1.	Diseño experimental .....	22
3.2.	Material vegetal .....	23
3.2.1.	Identificación de las muestras .....	24
3.3.	Adquisición de imágenes hiperespectrales (HSI).....	24
3.4.	Procesado de imágenes .....	26
3.4.1.	Formación del hipercubo .....	26
3.4.2.	Normalización .....	26
3.4.3.	Segmentación.....	27
3.4.4.	Unfolding.....	27
3.5.	Análisis quimiométrico: análisis multivariante .....	29
4.	RESULTADOS Y DISCUSIÓN .....	34
4.1.	Modelos PLS-DA para clasificar hojas de patata según los efectos de la fertilización .....	34
4.2.	Contraste de resultados con el método de referencia .....	40
4.3.	Validación externa de los modelos de calibración.....	42
5.	CONCLUSIONES .....	49
6.	BIBLIOGRAFÍA.....	50

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Métodos de detección/formación de imágenes hiperespectrales (HSI) (Wu y Sun, 2013)....	15
<b>Tabla 2.</b> Ventajas e inconvenientes del empleo de la técnica de imágenes hiperespectrales (Sun, 2010) .....	18
<b>Tabla 3.</b> Pretratamientos de los modelos de calibración desarrollados (MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada) .....	31
<b>Tabla 4.</b> Matriz de confusión binaria .....	33
<b>Tabla 5.</b> Características de los modelos de calibración desarrollados. Resaltados los modelos obtenidos más precisos (CC: Correctamente Clasificadas; CV: Validación Cruzada; LV: Variables Latentes; NONE: crudo; MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada) .....	35
<b>Tabla 6.</b> Resultados de sensibilidad, especificidad y error de los modelos de calibración desarrollados calculados para la clase “control”. Resaltado, para cada parámetro, los mejores resultados obtenidos (Cal: Calibración; CV: Validación Cruzada; NONE: crudo; MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada) .....	38
<b>Tabla 7.</b> Características del modelo de calibración generado para la clasificación de patatas en función del efecto de la variedad (CC: Correctamente Clasificadas; CV: Validación Cruzada; LV: Variables Latentes; MC: Mean Centering; SNV: Standard Normal Variate; SM: Smoothing) .....	39
<b>Tabla 8.</b> Resultados estadísticos descriptivos del análisis ANOVA de un factor (tratamiento control/déficit de nitrógeno) para la variable rendimiento (kg/ 10 plantas) .....	41
<b>Tabla 9.</b> Resultados del análisis ANOVA de un factor (tratamiento control/déficit de nitrógeno) para la variable rendimiento (kg/ 10 plantas).....	41
<b>Tabla 10.</b> Porcentaje de muestras correctamente clasificadas (%CC) en la validación externa de cada modelo de calibración desarrollado en el presente estudio. Resaltados los modelos más precisos obtenidos tras la validación externa. (NONE: crudo; MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada) .....	44
<b>Tabla 11.</b> Resultados de la validación externa del modelo de calibración número 6 y 13 (Pred: predicción; MC: Mean Centering; SNV: Standard Normal Variate; SM: Smoothing) .....	46

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Principales impactos del cambio climático en el sector agrícola en las principales regiones biogeográficas (EEA, 2019). .....	8
<b>Figura 2.</b> Hipercubo tridimensional con (3 filas) × (3 columnas) × (3 longitudes de onda) (Dorrepaal et al., 2016). .....	12
<b>Figura 3.</b> Cortes de longitud de onda de un hipercubo (Dorrepaal et al., 2016).....	12
<b>Figura 4.</b> Esquema de un sistema de adquisición de imágenes hiperespectrales (Castro, 2015). .....	13
<b>Figura 5.</b> Métodos de adquisición de imágenes hiperespectrales (Wu y Sun, 2013).....	15
<b>Figura 6.</b> Diagrama de flujo del proceso de análisis de imágenes hiperespectrales. ....	16
<b>Figura 7.</b> Representación gráfica del despliegue (unfolding) de un hipercubo (Dorrepaal et al., 2016) .....	17
<b>Figura 8.</b> Equipo de adquisición de imágenes hiperespectrales empleado: equipo real (a); diagrama esquemático (b) (Lopez Maestresalas, 2016) .....	24
<b>Figura 9.</b> Esquema del funcionamiento del espectrógrafo en el sistema de adquisición de imágenes hiperespectrales (Municio, 2011) .....	25
<b>Figura 10.</b> Espectro con ruido instrumental .....	28
<b>Figura 11.</b> Espectro sin ruido instrumental tras su eliminación .....	28
<b>Figura 12.</b> Diagrama de flujo del procedimiento del análisis quimiométrico de los datos espectrales .....	30
<b>Figura 13.</b> Representación del error cuadrático medio de calibración y validación cruzada en función de las variables latentes (LV) .....	32
<b>Figura 14.</b> Espectros de las muestras del grupo de calibración en crudo, sin aplicación de pretratamientos (NONE).....	36
<b>Figura 15.</b> Efecto de los pretratamientos en los espectros de las muestras del grupo de calibración: (MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada).....	37
<b>Figura 16.</b> Predicción en la validación cruzada (CV) de la clasificación de las muestras correspondientes con la variedad Mayka).....	40
<b>Figura 17.</b> Efecto de los pretratamientos en los espectros de las muestras del grupo de validación: a) NONE, b) SM y c) SM+SNV (NONE: Sin pretratamiento; SM: Smoothing; SNV: Standard Normal Variate).....	43
<b>Figura 18.</b> Representación gráfica de los resultados de predicción que ejecuta el modelo (a) número 6 y, (b) número 13, sobre la matriz de validación externa .....	45
<b>Figura 19.</b> Imagen química de las 48 muestras empleadas para la validación externa de los modelos PLS-DA desarrollados .....	47

## 1. INTRODUCCIÓN Y OBJETIVO

El cultivo de patata (*Solanum tuberosum* L.) es considerado, en términos de consumo humano, el tercer cultivo más importante del mundo (International Potato Center, 2017). Es cultivado en todos los continentes a excepción de la Antártida. En 2020 se produjeron a nivel mundial 359.071.403 toneladas de patatas, siendo China el mayor productor con una producción de 78.183.874 toneladas, seguido de India (51.300.000 toneladas) (FAO, 2020). La Unión Europea (UE) produjo 54 millones de toneladas de patatas en 2020 lo que supuso un aumento del 5,60% respecto a la producción de 2019 (Eurostat, 2020).

Gracias a las tendencias en el desarrollo tecnológico en las últimas décadas, en cuanto al procesamiento de este tubérculo, se ha producido una aceleración en el crecimiento de este segmento del mercado y una mejora en la alimentación de la población de los países en desarrollo y desarrollados (Loyola et al., 2010). Es por ello por lo que, el cultivo de patata es considerado fundamental tanto para asegurar la seguridad alimentaria como para contribuir en la reducción de la pobreza y mejora de la nutrición humana (López Maestresalas et al., 2021).

La patata es uno de los hidratos de carbono que más energía nos aportan debido, en parte, a su contenido en almidón. La patata recién cosechada, según la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), contiene un 80% de agua y un 20% de materia seca. Dentro del porcentaje de materia seca, el almidón representa entre el 60% y el 80% de esta (FAO, 2008). Sin embargo, estos porcentajes podrían verse alterados como consecuencia del cambio climático.

El cambio climático es uno de los mayores retos de la agricultura moderna al cual se enfrenta, entre otros muchos, el cultivo de la patata. El aumento de los gases efecto invernadero (GEI) es una de las principales causas del cambio climático. En este escenario de cambio se encuentra nuestro planeta Tierra. Los niveles de CO<sub>2</sub> en la atmósfera se han incrementado de forma exponencial en los últimos años provocando un incremento de los estreses abióticos sobre el cultivo de la patata. Este hecho está amenazando la sostenibilidad de su producción (George et al., 2017) debido, entre otras, a que las elevadas concentraciones de CO<sub>2</sub> en la atmósfera, a largo plazo, podrían alterar la toma y transporte de nutrientes, en especial, del nitrógeno (N) (Easlon & Bloom, 2013). La gran importancia de esto radica en que una deficiencia de nitrógeno se podría ver reflejada en una menor producción de almidón en la patata y, por ende, en una reducción de la calidad del tubérculo.

Por todo ello, resulta fundamental desarrollar una tecnología que permita identificar, de forma temprana, síntomas derivados del impacto del cambio climático en las plantas de patatas con el fin de poder anticiparse en campo a los efectos perjudiciales. Se plantea afrontar este reto a través del uso de tecnologías basadas en imagen ya que estas presentan grandes ventajas al ser capaces de adquirir datos de un número elevado de plantas, de forma rápida y no destructiva (López Maestresalas et al., 2021). La aplicación de imágenes hiperespectrales (HSI) en el rango infrarrojo cercano (NIR), en este caso, podría permitir el desarrollo de un método rápido, no destructivo y fiable para la caracterización del cultivo de patata (López-Maestresalas et al., 2016).

Teniendo en cuenta la importancia del cultivo de patata y la problemática actual relacionada con la situación de cambio global en la cual nos encontramos, el objetivo de este estudio es evaluar el potencial de la técnica HSI en el rango NIR como herramienta para identificar la respuesta, en hojas de plantas de patata, a la deficiencia de abonado nitrogenado (estrés abiótico controlado).

La puesta en práctica de esta técnica permitiría realizar una fertilización nitrogenada eficiente e identificar aquellos cultivares de patata más resistentes a la deficiencia de nitrógeno para incluirlos en programas de mejora genética futuros.

## 2. REVISIÓN BIBLIOGRÁFICA

### 2.1. Cambio Global

Objetivo 13: “Adoptar medidas urgentes para combatir el cambio climático y sus efectos” (ONU, 2021b). El objetivo 13, es uno de los 17 Objetivos de Desarrollo Sostenible (ODS) con los que cuenta la Agenda 2030 sobre el Desarrollo Sostenible. La Agenda 2030 fue aprobada por la ONU en 2015 como una oportunidad para que todos los países y sus sociedades emprendieran un nuevo camino con el que mejorar la vida de todos, sin dejar atrás a nadie. Tal es así que, entre los objetivos con los que cuenta la Agenda se incluyen desde la eliminación de la pobreza hasta el combate al cambio climático, la educación, la igualdad de la mujer, la defensa del medio ambiente o el diseño de nuestras ciudades (ONU, 2021b).

Entre otros, actualmente nuestro planeta se encuentra ante una situación de cambio global, especialmente, ante un cambio climático. Es por ello por lo que se incluyó en la Agenda 2030 sobre el Desarrollo Sostenible un objetivo encaminado hacia la acción por el clima, el Objetivo nº 13. En el artículo 1 de la Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC), se define el cambio climático como “cambio de clima atribuido directa o indirectamente a la actividad humana que altera la composición de la atmósfera global y que se suma a la variabilidad natural del clima observada durante períodos de tiempo comparables” (Solana, 1994). Así, la CMNUCC diferencia entre el cambio climático atribuible a las actividades humanas que alteran la composición atmosférica y la variabilidad climática atribuible a causas naturales. Cabe destacar que, en 1992, se adoptó la CMNUCC como consecuencia de la preocupación mundial existente sobre el calentamiento del planeta con el propósito de estabilizar la concentración de GEI en la atmósfera para evitar las alteraciones del sistema climático mundial provocadas por los seres humanos (Ibárcena Escudero y Scheelje Bravo, 2003).

Siguiendo la misma línea, la Comisión Europea especifica que los seres humanos influyen cada vez más en el clima y la temperatura del planeta Tierra como consecuencia de la quema de combustibles fósiles, la tala de bosques y la cría de ganado. Además, la Comisión Europea, coincidiendo con lo dictado por la CMNUCC, añade que las enormes cantidades de gases producidos por dichas actividades humanas se suman a los que se liberan de forma natural en la atmósfera, aumentando así el efecto invernadero y, por ende, el calentamiento global (Comisión Europea, 2021a).

#### 2.1.1. Causas y consecuencias del cambio climático

Se habla de cambio climático cuando el efecto invernadero se convierte en un fenómeno con consecuencias negativas para la vida planetaria. El efecto invernadero es conocido como un fenómeno natural por el que determinados gases que componen la atmósfera, los llamados gases de efecto invernadero (GEI), retienen parte de la energía solar reflejada por el suelo, absorbiéndola y transformándola en un movimiento molecular interno que produce un aumento de la temperatura de la Tierra. La presencia del efecto invernadero es necesaria para mantener la temperatura de nuestro planeta en unos valores habitables. Sin embargo, debido a la actividad antrópica, las concentraciones de algunos de los GEI están aumentando en la atmósfera provocando que el efecto invernadero pase de ser un gran aliado a ser un riesgo para nuestra supervivencia (Comisión Europea, 2021a). Cabe destacar el aumento de: dióxido de carbono (CO<sub>2</sub>), metano (CH<sub>4</sub>), óxido nitroso (N<sub>2</sub>O) y, gases fluorados. Según los datos arrojados en el informe resumen del *Inventario nacional de emisiones de gases de efecto invernadero* (serie 1990-2019) publicado en marzo de 2021, en el año 2019, el total de emisiones netas medidas en CO<sub>2</sub> equivalente (kt) de CO<sub>2</sub>, CH<sub>4</sub> y N<sub>2</sub>O fueron de 213.507,90, 38.644,60 y 18.583,50 respectivamente (MITECO, 2021b).

Son varias las causas que han intensificado dichas emisiones y, por tanto, el efecto invernadero, entre las que se encuentran:

- La combustión fósil, de biomasa y de residuos la cual produce, entre otros, CO<sub>2</sub> y N<sub>2</sub>O (Ibárcena Escudero y Scheelje Bravo, 2003).
- Los cambios en los usos del suelo como la tala y quema de bosques (deforestación) alterando como resultado el albedo superficial (Ibárcena Escudero y Scheelje Bravo, 2003).
- La liberación a la atmósfera del carbono almacenado en los árboles como consecuencia de la deforestación de los bosques, los cuales se caracterizan por ser los principales sumideros naturales de CO<sub>2</sub> (Comisión Europea, 2021a).
- El desarrollo de la ganadería. Actividad que está causando un aumento de la cantidad de CH<sub>4</sub> en la atmósfera debido a la fermentación entérica como consecuencia del proceso digestivo de los herbívoros (Ibárcena Escudero y Scheelje Bravo, 2003) y la descomposición en condiciones anaerobias del estiércol generado (Varnero Moreno, 2011).
- La falta de eficiencia en la fertilización nitrogenada que, como consecuencia, produce emisiones de N<sub>2</sub>O (Comisión Europea, 2021a).
- El tratamiento anaerobio de aguas residuales domésticas e industriales (Ibárcena Escudero y Scheelje Bravo, 2003).
- La fundición del aluminio, la fabricación de semiconductores y la transmisión y distribución de energía eléctrica, emitiendo los gases fluorados: PFCs y SF<sub>6</sub> (Ibárcena Escudero y Scheelje Bravo, 2003).

El aumento de la proporción de gases de efecto invernadero por encima de las concentraciones normales debido, entre otras, a las causas anteriormente citadas, ha provocado, tal y como se ha mencionado, que el efecto invernadero natural terrestre se multiplicara dando lugar a un fenómeno perjudicial. A pesar de que esta mayor cantidad de GEI en la atmósfera puede tener consecuencias diversas, los efectos de este hecho van principalmente dirigidos hacia un aumento de la temperatura global de la Tierra y un desencadenamiento de efectos en las corrientes marinas, movimientos atmosféricos y en las dinámicas terrestres.

En base al último informe del Grupo Intergubernamental de Expertos sobre el Cambio Climático, IPCC por sus siglas en inglés, publicado el 7 de agosto de 2021, se conoce que los científicos ya han observado cambios en el clima de nuestro planeta en todas las regiones y en el conjunto del sistema climático (IPCC, 2021). Pese a que una reducción sustancial y permanente de las emisiones de CO<sub>2</sub> y de otros GEI permitiría limitar el cambio climático, se estima que las temperaturas se estabilizarían en un periodo de entre 20-30 años (IPCC, 2021). Por el contrario, la mejora de la calidad del aire sería notoria rápidamente.

Con todo ello, se ha observado que el cambio climático afecta hoy en día de multitud de maneras diferentes a todas las regiones de la Tierra, no es cuestión únicamente del aumento de temperatura. Las distintas regiones experimentan diversos cambios los cuales se intensificarán si aumenta el calentamiento. Concretamente se habla de cambios en la humedad y la sequedad, los vientos, la nieve y el hielo, en las zonas costeras y los océanos (IPCC, 2021). A continuación, se mencionan diferentes cambios experimentados como consecuencia del cambio climático:

- Deshielo de las capas de hielo de los polos y glaciares y aumento del nivel del mar debido al incremento de la temperatura. Se prevé que la temperatura media mundial durante los próximos 20 años alcanzará o llegará a superar un calentamiento de 1,5 °C (IPCC, 2021).

- Inundaciones y erosión en las zonas costeras y de baja altitud como consecuencia del deshielo y del aumento del nivel del mar. A finales del siglo XXI los fenómenos relacionados con el nivel del mar extremo podrían registrarse con una frecuencia anual mientras que antiguamente se producían 1 vez cada 100 años (IPCC, 2021).
- Deterioro de la calidad del agua, inundaciones y posible progresiva disminución de los recursos hídricos como consecuencia del incremento de la frecuencia de las fuertes lluvias y otros fenómenos climáticos extremos.
- Peligros para la salud humana
  - Se ha registrado en ciertas regiones (noroeste de EE.UU. u oeste de Canadá, entre otras) un incremento de muertes producidas por las altas temperaturas (WMO, 2021).
  - Los cambios en el clima influyen sobre la frecuencia y la distribución a nivel global de ciertas enfermedades transmitidas por vectores, así como sobre la dinámica estacional e interanual de patógenos, vectores, hospedadores y reservorios (Rabadian-Diehl, 2017). A modo de ejemplo, cabe destacar la aparición del mosquito portador (*Ae. aegypti* y *Ae. albopictus*) del virus del Zika en Canadá el cual es propio del centro y sur de América (NCEZID, 2019).
- Riesgos para la naturaleza
  - Dificultad de adaptación de gran cantidad de especies, como el bambú, ante la rapidez del cambio climático y sus consiguientes efectos. Entre otros, los actuales hábitats del panda se están volviendo rápidamente más cálidos y secos impidiendo la adaptación del bambú y como consecuencia, afectando el desarrollo del panda (*A. melanoleuca*) (WWF, 2018).
  - Posible incremento del riesgo de extinción de ciertas especies como consecuencia del aumento descontrolado de las temperaturas medias globales (Urban, 2015). Se prevé una desaparición de casi un 8% de las especies actuales como consecuencia del cambio climático.
  - Afección de los ecosistemas marinos debido a los cambios en el océano, como su calentamiento y acidificación, el aumento de la frecuencia de las olas de calor marinas y, la reducción de los niveles de oxígeno (IPCC, 2021). Estos cambios están relacionados con la influencia del ser humano.
- Costes sociales y económicos
  - Pérdidas económicas directas debido a catástrofes naturales como las ocasionadas entre 1980 y 2011 debido a las inundaciones. Las pérdidas económicas alcanzaron los 90 mil millones de euros y más de 5,5 millones de personas fueron afectadas (Comisión Europea, 2021b).
  - Impactos negativos sobre los sectores dependientes del clima (temperatura y precipitaciones) como la agricultura, silvicultura, energía y el turismo (Comisión Europea, 2021b).
  - Repercusión en los sectores que dependen de los ecosistemas marinos como consecuencia del calentamiento, acidificación, aumento de la frecuencia de las olas de calor marinas y la reducción de O<sub>2</sub> de los océanos (IPCC, 2021).

A pesar de la situación de cambio global en la que se encuentra la Tierra y conocidas o estimadas las posibles consecuencias de este cambio, la Copresidenta del Grupo de Trabajo I del IPCC, Valérie Masson-Delmotte, da un halo de esperanza. Afirma que, gracias a las innovaciones y los avances en climatología, hoy en día tenemos una visión más clara del clima pasado, presente y futuro. Este hecho es esencial para entender hacia dónde nos dirigimos, qué se puede hacer y cómo podemos prepararnos.

En el informe del Grupo de Trabajo I del IPCC, *Cambio Climático 2021: Bases físicas*, aprobado el 6 de agosto de 2021, quedan reflejados los avances en climatología y las innovaciones mencionadas (IPCC, 2021).

### 2.1.2. Repercusión del Covid-19 en el cambio climático

Las restricciones de movimiento y las recesiones económicas derivadas de la pandemia provocaron una caída de entorno al 6% de las emisiones de gases de efecto invernadero en 2020 (ONU, 2021a). Sin embargo, tal y como afirma el Secretario General de la Organización Meteorológica Mundial (OMM), Petteri Taalas, la caída de las emisiones de GEI es lamentablemente una buena noticia a corto plazo, en referencia a una caída de entre 5,5-5,7% en los niveles de CO<sub>2</sub> debido a la pandemia (ONU, 2020b). Así mismo, el jefe de la agencia meteorológica de la ONU especifica que el cambio climático no se va a pausar ya que una vez que la economía mundial comience a recuperarse de la crisis sanitaria se espera que las emisiones vuelvan a la normalidad o incluso aumenten (ONU, 2020b).

Según el informe de la OMM *Global Climate in 2015-2019*, los niveles de dióxido de carbono fueron un 18% más altos entre 2015 y 2019 que los 5 años anteriores (WMO, 2020a). En dicho informe se recogieron los datos que indicaban que las concentraciones de CO<sub>2</sub> estaban en camino de alcanzar o incluso superar las 410 ppm para finales de 2019. Este hecho se confirmó meses más tarde y se plasmó en el *Boletín de Gases de Efecto Invernadero* de la OMM. A finales de 2019, el promedio mundial anual de los niveles de dióxido de carbono superó el umbral significativo de 410 ppm, concretamente se alcanzaron 410,5 ppm (WMO, 2020b). Pese a la caída del 6% en 2020, se conoce que el CO<sub>2</sub> permanece en la atmósfera y los océanos durante siglos. Es decir, la Tierra está comprometida con el cambio climático continuo independientemente de cualquier caída temporal en las emisiones debido a la pandemia de Covid-19.

Cierto es que se ha contemplado una mejora de la calidad del aire en las principales ciudades y regiones industrializadas en varias partes del mundo. Esto se ha relacionado con la disminución de los niveles de contaminantes atmosféricos comunes de los escapes de los automóviles y la energía de los combustibles fósiles, como las partículas de N<sub>2</sub>O (ONU, 2020b). Sin embargo, tal y como afirmó Taalas, la caída de las emisiones de GEI es lamentablemente una buena noticia a corto plazo.

Es irrefutable que la reducción de las emisiones de dióxido de carbono debido a la pandemia es solamente un impacto positivo visible temporal. Pese a ello, siguiendo la afirmación de Inger Andersen, directora del Programa de las Naciones Unidas para el Medio Ambiente (PNUMA), *“la crisis actual es una oportunidad para llevar a cabo un cambio profundo hacia una economía más sostenible que funcione tanto para el planeta como para las personas que habitan en él”* (ONU, 2020c). Ante este hecho, las Naciones Unidas (ONU) instan a los Gobiernos a aprovechar la oportunidad de “reconstruir mejor” mediante la creación de sociedades más sostenibles, resilientes e inclusivas (ONU, 2020a). Siguiendo esta línea, el Secretario General de la ONU ha propuesto las siguientes 6 medidas favorables para el clima las cuales pueden adoptar los Gobiernos una vez que se dediquen a reconstruir sus economías y sociedades (ONU, 2021a):

1. Transición verde: las inversiones deben acelerar el proceso de reducción de emisiones de carbono (descarbonización) de todos los aspectos de nuestra economía.
2. Empleos verdes y crecimiento sostenible e inclusivo.
3. Economía verde: hacer que las sociedades y los pueblos sean más resilientes a través de una transición justa para todos, sin dejar a nadie atrás.
4. Invertir en soluciones sostenibles: eliminar los subsidios a los combustibles fósiles y obligar a los contaminadores a pagar por su contaminación.

5. Afrontar todos los riesgos climáticos.
6. Cooperación: ningún país puede triunfar por sí solo.

Englobando tanto las causas y consecuencias del cambio climático (véase “2.1.1. *Causas y consecuencias del cambio climático*”) como la respuesta del cambio climático al Covid-19 (véase “2.1.2. *Repercusión del Covid-19 en el cambio climático*”), los planes de recuperación posteriores a la actual crisis sanitaria deben dar pie a cambios sistémicos a largo plazo con el objetivo de que cambien la trayectoria de los niveles de dióxido de carbono, entre otros GEI, en la atmósfera (ONU, 2021a).

## 2.2. Interrelación del cambio climático y la agricultura

La agricultura contribuye al cambio climático y se ve afectada por el cambio climático. Es cierto que, en base a la contribución del sector agrícola al cambio climático, para hacerle frente, la UE necesita reducir sus emisiones de gases de efecto invernadero que tienen origen en la agricultura, así como adaptar el sistema de producción de alimentos (EEA, 2021). Sin embargo, el cambio climático es solo una de las muchas presiones que existen sobre la agricultura. Ya que este sector se ve presionado por la creciente demanda mundial y la competencia por los recursos, la producción y el consumo de alimentos de la UE deberían verse en un contexto más amplio el cual vincule la agricultura, la energía y la seguridad alimentaria (EEA, 2021).

### 2.2.1. Efectos de la agricultura en el cambio climático

Los especialistas de la Agencia Europea de Medio Ambiente, EEA de sus siglas en inglés, afirman que, “antes de llegar a nuestros platos, nuestra comida se produce, almacena, procesa, empaqueta, transporta, prepara y sirve” (EEA, 2021). En cada etapa que ocurre durante el aprovisionamiento de alimentos, se liberan GEI a la atmósfera. Concretamente, la agricultura libera cantidades significativas de CH<sub>4</sub> y N<sub>2</sub>O, dos poderosos gases de efecto invernadero.

En 2012 este sector representó el 10% de las emisiones totales de gases de efecto invernadero de la Unión Europea (UE). Sin embargo, cabe destacar que la contribución de este sector al cambio climático disminuyó en la UE entre 1990 y 2012 en un 24%. Este hecho no se repitió a nivel mundial ya que entre 2001 y 2011 las emisiones globales de la producción agrícola y ganadera aumentaron un 14%. Principalmente fue debido a un aumento de la producción agrícola total como consecuencia del aumento de la demanda mundial de alimentos y los cambios de patrones de consumo de alimentos al incrementar los ingresos en algunos países en desarrollo (EEA, 2021).

Según los datos aportados por el Ministerio para la Transición Ecológica y el Reto Demográfico los cuales se recogen en el *Avance del Inventario de Emisiones de Gases de Efecto Invernadero correspondientes al año 2020* (MITECO, 2021a), las emisiones brutas de GEI a nivel nacional en 2020 se estiman en 271,5 millones de toneladas de CO<sub>2</sub>-eq. Esto supone una disminución global de las emisiones de CO<sub>2</sub>-eq en 2020 respecto al 2019 del 13,7 %. Sin embargo, analizando el peso de las emisiones por sector, en lo que a la agricultura y ganadería en conjunto respecta, este supone un 14,1 % en el peso global de las emisiones de gases efecto invernadero en 2020, es decir, un 1,2 % más que el año anterior. Este aumento se debe principalmente al ligero crecimiento de las cabañas ganaderas (64,8 % de las emisiones de este sector) ya que se vieron aumentadas sus emisiones en un 0,8 % debido a las procedentes de la gestión de estiércol y de la fermentación entérica en menor medida. Además, la variación en las emisiones debidas a los cultivos también influyó en el aumento de las emisiones del sector. Estas aumentaron un 2 %, fundamentalmente por el aumento de las emisiones de N<sub>2</sub>O derivadas de la gestión de suelos agrícolas (1,5 % respecto al año anterior, debido al uso de fertilizantes inorgánicos) y de emisiones de CO<sub>2</sub> derivadas de la aplicación de urea (21 % respecto al año anterior).

En este caso, la estimación de las emisiones de GEI procedentes del sector se realizan siguiendo el manual que suministra la FAO a los países miembros (FAO, 2015b). Este manual es una herramienta y metodología para ayudar a identificar, construir y acceder al conjunto mínimo de datos de actividad necesarios para estimar los GEI de cada sector. En la estimación de las emisiones GEI procedentes del sector agricultura, silvicultura y otros usos de la tierra se incluyen (FAO, 2015b):

- Emisiones de CO<sub>2</sub> resultantes de las variaciones de las reservas de carbono en la biomasa, materia orgánica muerta, materia orgánica del suelo (suelos orgánicos y minerales) y, productos de madera recolectada.
- CO<sub>2</sub> procedente de los suelos orgánicos cultivados.
- Emisiones distintas del CO<sub>2</sub> provenientes de incendios.
- Emisiones de CH<sub>4</sub> procedentes del cultivo del arroz
- Emisiones de N<sub>2</sub>O de todos los suelos gestionados.
- Emisiones de CO<sub>2</sub> asociadas con la aplicación de cal y urea a los suelos gestionados.
- Emisiones de CH<sub>4</sub> que provienen de la fermentación entérica del ganado.
- Emisiones de CH<sub>4</sub> y N<sub>2</sub>O procedentes de los sistemas de gestión del estiércol.

Teniendo en cuenta los datos arrojados y la importancia de los alimentos en nuestras vidas, alcanzar una reducción mayor de las emisiones de GEI del sector agrícola sigue siendo un reto. No obstante, sigue existiendo la posibilidad de reducir las emisiones vinculadas con la producción de alimentos a través de la integración de técnicas innovadoras en los métodos de producción. La captura de CH<sub>4</sub> del estiércol, un uso más eficiente de fertilizantes (agricultura de precisión) y, un aumento en la eficiencia de producción de carne y lácteos, es decir, disminuir las emisiones por unidad de alimento producido, puede ayudar a superar el desafío (EEA, 2021). Así mismo, cambiar las dietas, como comer menos carne y reducir el desperdicio de alimentos, contribuiría a reducciones adicionales (EEA, 2020a).

### 2.2.2. Efectos del cambio climático en la agricultura

Los cultivos necesitan tierra, agua, luz solar y calor adecuados para desarrollarse de forma óptima. Es por ello por lo que, aunque todos los sectores económicos están siendo afectados por el cambio climático, debido a la dependencia de la producción agrícola de las condiciones climáticas y meteorológicas, el sector agrícola es uno de los más vulnerables frente al cambio climático (EEA, 2020b).

En el informe *Adaptación al cambio climático en el sector agrícola en Europa* publicado por la Agencia Europea de Medio Ambiente (EEA) en 2019, se examina la incidencia del cambio climático en el sector agrícola y presenta unas perspectivas para los próximos años (EEA, 2019). En base a dicho informe y la información recogida en el documento generado por la FAO en 2015 que lleva por título *Cambio climático y seguridad alimentaria: riesgos y respuestas*, se puede afirmar que los cambios en la temperatura media y las precipitaciones, junto a los fenómenos meteorológicos extremos ya están influyendo en los rendimientos de los cultivos y productividad del ganado en Europa (FAO, 2015a).

Especifican que la influencia del cambio climático en el sector agrícola no será igual en todas las regiones del planeta Tierra. De hecho, los impactos serán positivos o negativos en función de las especies y las regiones geográficas. Además, la influencia dependerá de una serie de factores (y su combinación), como los impactos físicos, los cambios de los agroecosistemas y de la respuesta adaptativa de la sociedad (EEA, 2016). En lo que a los impactos físicos se refiere, cabe destacar los determinados cambios de temperatura, de los patrones de precipitación y de la concentración de CO<sub>2</sub> en la atmósfera. Los cambios en los agroecosistemas están relacionados con las posibles pérdidas de polinizadores y aumento de ciertas plagas y enfermedades.

En la Figura 1 se muestran los principales impactos del cambio climático en el sector agrícola en las principales regiones biogeográficas en Europa (EEA, 2019).

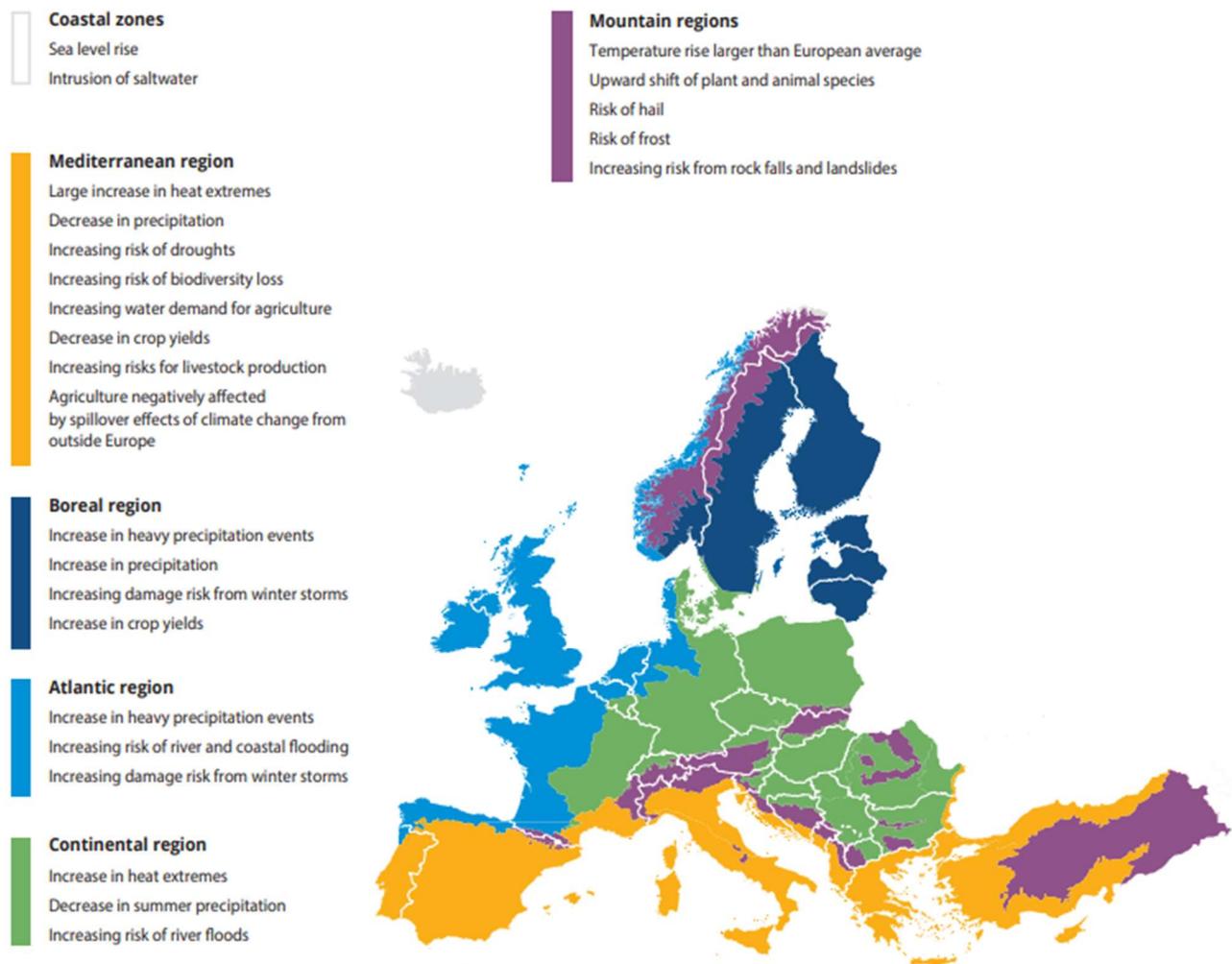


Figura 1. Principales impactos del cambio climático en el sector agrícola en las principales regiones biogeográficas (EEA, 2019).

El experto de la EEA en impacto y adaptación al cambio climático, Blaz Kurnik, recalca que el impacto en los rendimientos de los cultivos y la productividad ganadera puede conducir al abandono de tierras agrícolas desfavorecidas en partes del sur de Europa (EEA, 2020b). En general, en el norte de Europa la productividad agrícola podría aumentar debido a una temporada de crecimiento más larga y una extensión del período libre de heladas. Estos hechos podrían permitir la viabilidad de nuevos cultivos. Sin embargo, en el sur de Europa se espera que los eventos de calor extremo y la reducción de las precipitaciones y la disponibilidad de agua obstaculicen la productividad de los cultivos.

Para el caso de Europa, a continuación, se citan los principales impactos del cambio climático en la agricultura (EEA, 2019):

- Temporada de crecimiento y fenología de los cultivos
  - El aumento de la duración del ciclo vegetativo térmico ha hecho que se amplíen hacia el norte las zonas aptas para varios cultivos.
  - Se han observado cambios en la fenología de los cultivos, como el adelanto de las fechas de floración y cosecha en los cereales.

- Demanda de agua
  - El aumento previsto de la temperatura provocará un incremento de las tasas de evapotranspiración, aumentando así las necesidades de agua de los cultivos. Se espera que sea más aguda en el sur de Europa.
- Productividad de los cultivos
  - Se espera que la productividad de los cultivos aumente en el norte de Europa, como resultado de un alargamiento de la temporada de crecimiento. Por el contrario, se espera una disminución en el sur.
  - La disminución de la productividad de los cultivos en el sur será causada por una mayor velocidad de desarrollo de estos, con los consiguientes efectos negativos, especialmente en el llenado de los granos.
  - Los fenómenos meteorológicos y climáticos extremos pueden reducir considerablemente el rendimiento de algunos cultivos. Se prevé que este hecho aumente el riesgo de pérdidas de cosechas, con el consiguiente aumento de los precios de los alimentos y la reducción de la seguridad alimentaria.
  - Es probable que el cambio climático amplíe la actividad estacional de plagas y enfermedades y los riesgos asociados a estos efectos.

#### 2.2.2.1. Efecto de los GEI en la agricultura

Además del aumento de la temperatura media global, el cambio de los patrones de precipitación y de la alteración de la frecuencia de los fenómenos meteorológicos extremos, el incremento de la concentración de CO<sub>2</sub> en la atmósfera, es un importante impacto físico asociado con el cambio climático que tiene efecto sobre la agricultura. Pese a que se han realizado numerosos estudios sobre la influencia del cambio meteorológico (variación de temperatura y precipitaciones) en la agricultura, las evaluaciones del impacto del elevado dióxido de carbono se han centrado principalmente en los rendimientos de los cultivos (NASA, 2016).

Gracias al proceso de fotosíntesis llevado a cabo por las plantas, estas producen materia orgánica a partir de CO<sub>2</sub> empleando la luz solar como fuente de energía. Es por ello por lo que las especies vegetales son consideradas sumideros de carbono siendo esto fundamental en la lucha contra el cambio climático. Es un hecho que la vegetación necesita dióxido de carbono para desarrollarse y cumplir su ciclo vital de forma óptima. Sin embargo, se estima que el efecto beneficioso de un incremento de las concentraciones de CO<sub>2</sub> será limitado.

Un estudio de la NASA revela que el aumento de los niveles de dióxido de carbono ayudará y perjudicará los cultivos. Por un lado, según los resultados publicados en la revista *Nature climate Change* (Deryng et al., 2016), las concentraciones elevadas de CO<sub>2</sub> en la atmósfera pueden aumentar la eficiencia del uso del agua en los cultivos y mitigar considerablemente las pérdidas de rendimiento debido al cambio climático. Sin embargo, se prevé que dichos niveles del GEI tengan efectos perjudiciales sobre la nutrición de los cultivos. Tal y como afirmó la autora principal y científica climática del Instituto Goddard de Estudios Espaciales (GISS) de la Nasa, Delphine Deryng, los desequilibrios entre el nitrógeno y el carbono en los tejidos de los cultivos podrían conducir a una menor cantidad de nutrientes como hierro, zinc, junto con una reducción en el contenido de proteínas (NASA, 2016). Estudios revelan que el nitrógeno por unidad de masa foliar disminuye en promedio un 13% (Ainsworth y Long, 2005).

De este modo, han demostrado que el incremento de las [CO<sub>2</sub>] afectan a los cultivos de dos maneras importantes: (1) aumentan el rendimiento de los cultivos al aumentar la tasa de fotosíntesis, lo que estimula el crecimiento y, (2) reducen la cantidad de agua que los cultivos pierden por transpiración.

Esto es fruto de la relación con el proceso de transpiración de las plantas. Las plantas transpiran a través de sus hojas las cuales contienen poros diminutos denominados estomas que se abren y recolectan moléculas de CO<sub>2</sub> para la fotosíntesis. Durante ese proceso, liberan vapor de agua. Por lo tanto, las plantas regulan el grado de apertura estomática como un compromiso entre los objetivos de mantener altas tasas de fotosíntesis y bajas tasas de pérdida de agua. A medida que aumentan las concentraciones de dióxido de carbono, los estomas no se abren tanto y, por ende, los niveles de transpiración descienden. Este hecho da como resultado un aumento de la eficiencia del uso del agua.

Pese a que la reducción de la conductancia estomática debida al cierre parcial de los estomas beneficia al cultivo al aumentar la eficiencia del uso del agua, también provoca una disminución de la absorción del agua. El hecho de que las plantas absorban menos agua significa que también se reduce la absorción de los minerales esenciales del suelo como el nitrógeno. Junto a este factor, la dilución del nitrógeno a partir del aumento de las concentraciones de carbohidratos (Taub y Wang, 2008) y la disminución de la tasa de asimilación de nitratos (J., Martin, Rubio, y B., 2010) en compuesto orgánicos provoca la citada reducción de nitrógeno en los tejidos y el consiguiente desequilibrio con el carbono.

En resumen, la evidencia actual sugiere que [CO<sub>2</sub>] atmosféricas pronosticadas para los próximos años tendrán importantes implicaciones para la fisiología y el crecimiento de las plantas. Bajo niveles elevados de CO<sub>2</sub>, la mayoría de las especies vegetales muestran:

- tasas más altas de fotosíntesis,
- mayor crecimiento vegetativo,
- menor uso de agua y,
- menores concentraciones tisulares de nitrógeno y proteínas.

Así, es probable que el aumento del CO<sub>2</sub> durante el próximo siglo afecte tanto a la producción agrícola como a la calidad de los alimentos. La reducción en la calidad de los alimentos está relacionada con la menor concentración de proteínas en los tejidos vegetales (vinculado con el estado del nitrógeno de la planta). Esto a su vez repercutirá en importantes especies de los niveles tróficos más altos, incluso se estima que se produzcan efectos sobre la nutrición humana (Taub, 2010).

### **2.3. Imágenes Hiperespectrales (HSI)**

La tecnología de imágenes hiperespectrales, HSI de sus siglas en inglés, surge de la integración de las principales ventajas de las técnicas ópticas sin contacto como la espectroscopia y las imágenes. Ambas técnicas se caracterizan por ser muy poderosas para la inspección en línea de productos agrícolas y alimentarios con la finalidad de garantizar su calidad y seguridad.

Por un lado, la espectroscopia permite determinar las cualidades esenciales de los productos objeto de estudio basándose en la medición de las propiedades ópticas. Sin embargo, esta técnica no tiene la capacidad de facilitar la información respecto a la distribución espacial de rasgos en productos alimenticios (Wu y Sun, 2013).

Por otro lado, a través de un sistema de imágenes convencional, concretamente con visión por computadora, se pueden medir las características externas de los productos. No obstante, debido a que trabaja en longitudes de onda visibles en forma de imágenes monocromáticas o en color, este sistema no tiene la capacidad de inspeccionar muestras con colores similares ni de clasificar objetivos complejos, predecir componentes químicos o detectar defectos no visibles a simple vista (Wu y Sun, 2013).

Es por ello por lo que combinando las principales ventajas de ambas técnicas ópticas sin contacto se puede obtener simultáneamente información espectral y espacial de los productos agrícolas y alimentarios analizados (Wu y Sun, 2013). En definitiva, la técnica de HSI se puede emplear para la predicción cuantitativa de las propiedades químicas y físicas inherentes de la muestra, así como su distribución espacial de forma rápida y no destructiva. Concretamente en la industria agroalimentaria es común emplear la técnica de imágenes hiperespectrales (HSI) en el rango infrarrojo cercano (NIR).

En lo que a la espectroscopia de infrarrojo cercano (NIRS) se refiere, cabe mencionar que el infrarrojo cercano está formado por el conjunto de longitudes de onda ( $\lambda$ ) comprendidas entre 780-2.350 nanómetros (nm) (Amigo, Babamoradi, y Elcoroaristizabal, 2015).

### 2.3.1. Base teórica

Una imagen hiperespectral es un gran conjunto de datos en el que cada píxel corresponde a un espectro. Esto proporciona detalles de alta calidad de la superficie de la muestra (Dorrepaal, Malegori, y Gowen, 2016). Las imágenes HSI se caracterizan por proporcionar información espectral y espacial, permitiendo así la obtención de información tanto cualitativa como cuantitativa de la muestra objeto de análisis. Tal y como se ha mencionado anteriormente, la técnica de imágenes hiperespectrales en el rango del infrarrojo cercano brinda una potente herramienta para analizar de forma no destructiva y rápida diferentes materiales. Mediante el conocimiento previo de su información espectral, esta técnica permite supervisar en tiempo real la información espacial del material objeto de estudio.

Las imágenes hiperespectrales están constituidas por 2 dimensiones espaciales y 1 dimensión espectral. Los componentes espaciales son  $I \times J$ , siendo  $I$  las columnas y  $J$  las filas. Estos componentes forman una matriz en la que cada elemento de la matriz es considerado como un píxel de la imagen. Las intensidades reflejadas por cada píxel en las diferentes longitudes de onda medidas ( $\lambda$ ) constituyen la dimensión espectral. Es decir, se habla de una matriz cuyos índices de columna y fila ( $I \times J$ ) identifican un punto en la imagen y cuyo valor coincide con el nivel de intensidad de la luz de ese punto (Grahm y Geladi, 2007). Por lo tanto, la matriz tiene un perfil tridimensional de píxeles los cuales cada uno proporciona información espectral concreta permitiendo así obtener un espectro completo de cada uno de ellos (Mollazade, Omid, Tab, y Mohtasebi, 2012).

En definitiva, la información en una HSI se organiza en una matriz tridimensional denominada hipercubo (Figura 2). La información proporcionada del material analizado puede ser desde observaciones físicas y geométricas (tamaño, orientación, forma, color y textura), hasta información química/molecular (contenido de agua, grasa, proteínas) (Sun, 2010). A pesar de que los hipercubos pueden representar más de 100 longitudes de onda, en la representación gráfica mostrada en la Figura 2 sólo se detallan 3  $\lambda$  para facilitar la interpretación (Dorrepaal et al., 2016). Así mismo, en la Figura 3 se muestra cómo, en una determinada longitud de onda del hipercubo, la matriz puede conceptualizarse como un corte bidimensional del hipercubo (Dorrepaal et al., 2016). Cada corte puede visualizarse como una imagen que muestra las intensidades espaciales relativas de absorbancia/reflexión en esa longitud de onda.

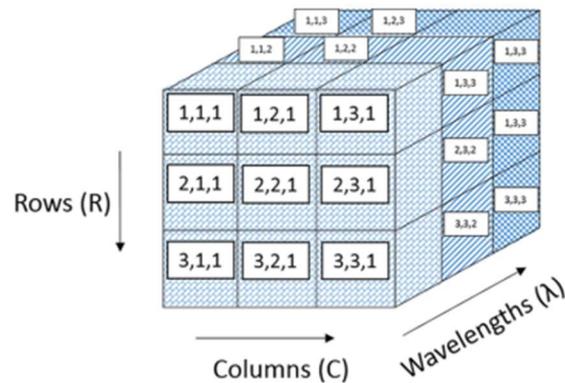


Figura 2. Hipercono tridimensional con (3 filas) × (3 columnas) × (3 longitudes de onda) (Dorrepaal et al., 2016).

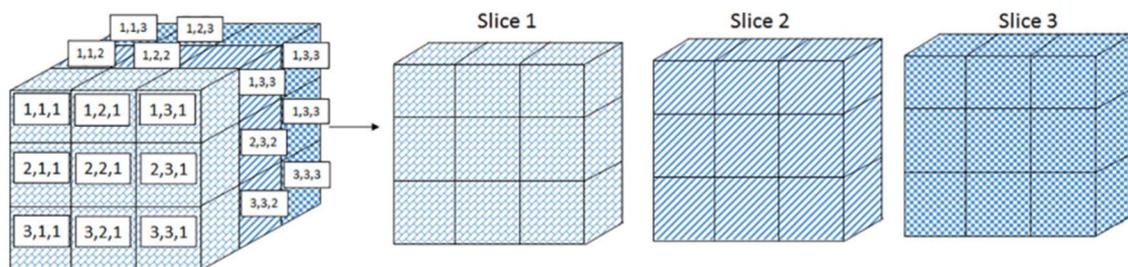


Figura 3. Cortes de longitud de onda de un hipercono (Dorrepaal et al., 2016).

### 2.3.2. Elementos del sistema de imágenes hiperespectrales

Los componentes o instrumentación del sistema HSI son esenciales para adquirir imágenes hiperespectrales fiables y de alta calidad. La selección de los elementos de los instrumentos, el diseño de su configuración y la calibración requieren una buena comprensión de la configuración y calibración del sistema HSI (Wu y Sun, 2013). Un sistema de imágenes hiperespectrales consiste en tres elementos principales: (1) fuente de luz, (2) dispositivo de dispersión de longitud de onda y, (3) detectores de área (Elmasry, Kamruzzaman, Sun, y Allen, 2012; Wu y Sun, 2013). En la Figura 4 se muestra el esquema de los elementos que componen un sistema de adquisición de imágenes hiperespectrales (Castro, 2015).

- Fuente de luz: tienen la función de generar luz como portador de información para excitar o iluminar el objetivo siendo una parte fundamental de los sistemas de inspección óptica. Las fuentes de luz típicas utilizadas en los sistemas de HSI son 4: diodos emisores de luz, láseres, fuentes de luz sintonizables y lámparas halógenas. Estas últimas son las más empleadas en las regiones espectrales del visible (VIS) y el infrarrojo cercano (NIR) (Wu y Sun, 2013). Se caracterizan por ofrecer un alto grado de direccionalidad dando lugar a variaciones espaciales en la intensidad de la fuente de luz en el plano de la escena. Por ello, para conseguir una iluminación uniforme del objeto, es importante que la fuente de luz se coloque en una disposición en la que se proporcione iluminación difusa (Gómez-Sanchis et al., 2008).
- Dispositivo de dispersión de longitud de onda: dispersan la luz provista, en este caso por la lámpara halógena, en diferentes longitudes de onda. Entre los dispositivos más comunes se encuentran los espectrofotómetros de imágenes, los filtros sintonizables de cristal-líquido LCTF (Liquid Crystal Tunable Filter) y, filtros sintonizables acústico-ópticos AOTF (Acoustic-Optic Tunable Filters) (Wu y Sun, 2013).

- Detectores de área: cuantifican la intensidad de la luz adquirida convirtiendo los fotones incidentes en electrones. Es decir, cuantifican de forma simultánea la información espacial y espectral en forma de luz tras pasar por el dispositivo de dispersión de longitud de onda (Wu y Sun, 2013). Para convertir la energía de la radiación en señales eléctricas se puede emplear o un dispositivo de carga acoplada CCD (Coupled Charged Device) o un semiconductor complementario de óxido metálico CMOS (Complementary Metal Oxide Semiconductor) (Elmasry, Kamruzzaman, et al., 2012).

A parte de estos 3 elementos principales, según el dispositivo de dispersión de longitud de onda empleado, cabe la posibilidad de que sea necesario el uso de una plataforma con la finalidad de que desplace la muestra a una velocidad constante mientras el detector de área (cámara) realiza la captura de la escena (Gowen, O'Donnell, Cullen, Downey, y Frias, 2007). Se requiere desplazamiento de la muestra en el caso que se emplee un espectrofotómetro de imagen como dispositivo de dispersión de longitud de onda (Castro, 2015).

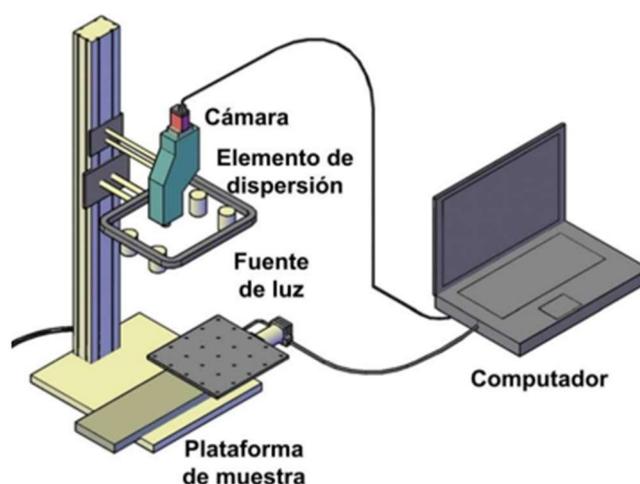


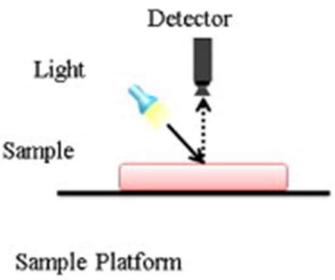
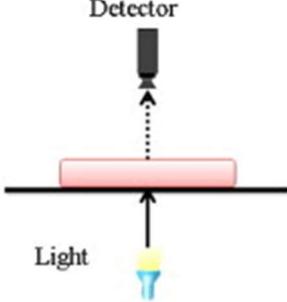
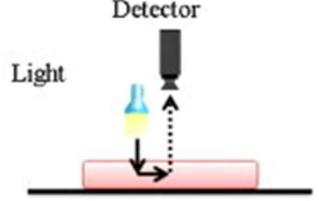
Figura 4. Esquema de un sistema de adquisición de imágenes hiperespectrales (Castro, 2015).

### 2.3.3. Métodos de detección y adquisición de imágenes hiperespectrales

La matriz tridimensional o hipercubo, dónde se almacena la información de la muestra analizada, se construye de acuerdo a los diferentes métodos de detección y adquisición (escaneo) de imágenes hiperespectrales. Tal y como se ha mencionado en el punto “2.3.2. Elementos del sistema de imágenes hiperespectrales”, dichos métodos dependen a su vez de la disposición y de los componentes constituyentes del sistema de HSI.

En lo que a los métodos de detección y formación de HSI respecta, los modos más habituales son: (1) reflectancia, (2) transmisión y, (3) fluorescencia. Lo que difiere de cada uno de los tres métodos es la posición de la fuente de luz y el detector de área (Tabla 1) (Castro, 2015; Gowen et al., 2007; Wu y Sun, 2013). El primer método, reflectancia, generalmente se emplea para obtener características de calidad externas de la muestra como tamaño, forma, color, textura superficial, etc. La transmisión se utiliza habitualmente para determinar la concentración de los componentes y defectos internos de muestras relativamente transparentes como pescado, frutas y verduras. El último y tercer método, la fluorescencia o interacción, se emplea para detectar información más profunda de la muestra. De todos ellos, el método más común de detección de imágenes hiperespectrales es la reflectancia.

Tabla 1. Métodos de detección/formación de imágenes hiperespectrales (HSI) (Wu y Sun, 2013)

Reflectancia	Transmisión	Fluorescencia/Interacción
El sensor captura la luz reflejada desde el mismo lado de donde se ilumina la muestra.	El sensor captura la luz transmitida a través de la muestra la cual lleva información interna.	El sensor, la fuente de luz y la muestra, se localizan en el mismo lado y de forma paralela entre sí.
		

En relación con los métodos de adquisición de imágenes hiperespectrales, existen 4 modos de configuraciones convencionales diferentes: (1) escaneo del área, (2) escaneo por puntos, (3) escaneo por líneas y, (4) escaneo en una única toma o single shot (Castro, 2015; ElMasry y Sun, 2010; Lopez Maestresalas, 2016; Wu y Sun, 2013). Debido a que estos métodos sólo capturan uno o dos dimensiones del subconjunto de hipercubos, excepto el escaneo en una única toma, precisan de un escaneo para obtener los componentes restantes. A continuación, se realiza una descripción de cada método y se recoge una representación de cada uno de ellos en la Figura 5:

- Escaneo del área: se obtiene una imagen monocromática en dos dimensiones con información espacial completa en una sola longitud de onda. Se requiere generar por cada longitud de onda una imagen. Por ello, es necesario un filtro para ajustar la longitud de onda deseada en cada momento. Sin embargo, cabe destacar que no es necesario mover la muestra porque el campo de visión de la imagen se mantiene fijo. La información del hipercubo resultante se almacena en formato BSQ (banda secuencial).
- Escaneo por puntos: mide y proporciona el espectro de un determinado punto al escanearlo en un píxel. La muestra debe moverse posteriormente para medir el espectro de otro punto diferente. El proceso se repite sucesivamente para obtener el cubo hiperespectral completo. El formato de salida, es decir, la información del hipercubo se almacena en formato BIP (banda intercalada por píxel).
- Escaneo por líneas: se mide y registra una línea completa de una imagen con información espectral de cada píxel espacial de la línea gracias a que la cámara posee una fila de detectores. A diferencia del escaneo por puntos, este método mide de una vez una línea de píxeles. El cubo hiperespectral completo se obtiene desplazando la muestra y el formato de la información de salida es BIL (banda intercalada por línea).
- Escaneo en una única toma: registra tanto la información espacial como la espectral empleando un detector de gran superficie con una sola exposición para capturar las imágenes. Aunque esto hace que sea un método muy atractivo cuando se requieren imágenes hiperespectrales rápidas, todavía se encuentra en una fase temprana de desarrollo y tiene resoluciones limitadas para la dimensión espacial y rangos estrechos para la dimensión espectral.

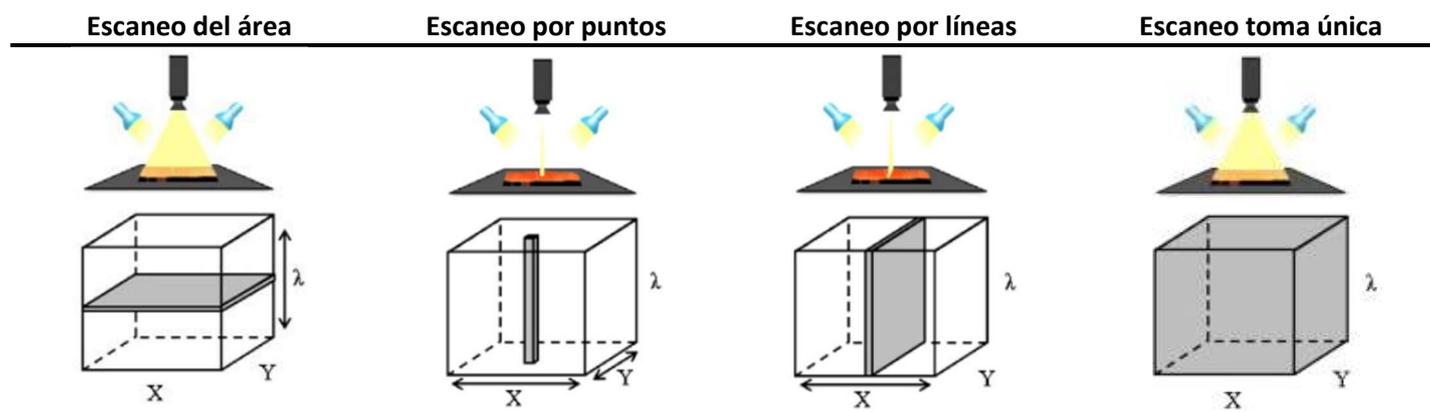


Figura 5. Métodos de adquisición de imágenes hiperespectrales (Wu y Sun, 2013)

#### 2.3.4. Etapas del proceso de análisis

En cuanto al método de detección/formación de HSI se refiere, cabe destacar que, generalmente, se escoge reflectancia. Así, tras seleccionar el método de detección de las imágenes hiperespectrales y, elegir la configuración y la iluminación, se adquiere la imagen de una muestra que se almacena en varios archivos (.png). El procedimiento de adquisición y análisis de imágenes hiperespectrales se puede dividir en dos etapas, (1) adquisición y preprocesado, de acuerdo con los métodos de adquisición detallados en el apartado "2.3.3. Métodos de detección y adquisición de imágenes hiperespectrales" y, (2) modelización o quimiometría. Como resultado de la primera etapa se obtiene la información espectral y, a través de las técnicas de análisis estadístico multivariante utilizadas en la segunda etapa, se correlacionan los parámetros analizados con dicha información espectral (Castro, 2015; Wu y Sun, 2013).

En la Figura 6, se ilustra el procedimiento de adquisición y análisis de imágenes hiperespectrales. Además, mientras en el presente apartado se detalla la primera etapa, adquisición y preprocesado, la segunda etapa queda explicada en el apartado "2.4. Quimiometría".

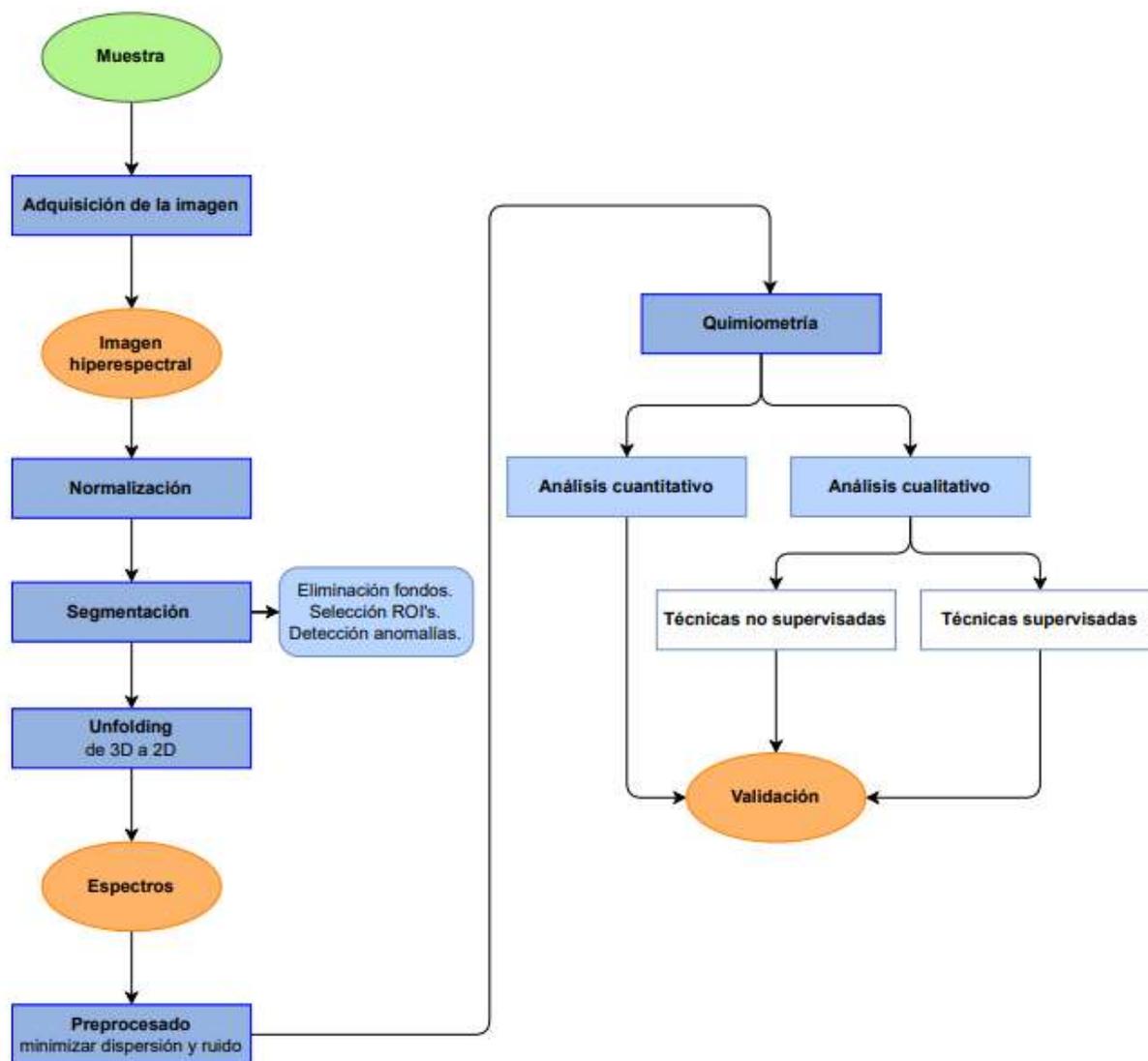


Figura 6. Diagrama de flujo del proceso de análisis de imágenes hiperespectrales.

La primera etapa (adquisición y preprocesado) en la que se divide el procedimiento de adquisición y análisis de HSI, permite tanto obtener y comprimir la información para reducir el almacenamiento requerido, como obtener una imagen hiperespectral corregida. Para la extracción de información a partir de la HSI es preciso seguir los siguientes pasos constituyentes de esta primera etapa: normalización, segmentación (eliminación de fondos, selección de regiones de interés (ROI's), detección de anomalías), unfolding y preprocesado. Tras la realización de dichos procesos la información espectral de la muestra será apta para iniciar el análisis quimiométrico (análisis multivariante).

Con la finalidad de normalizar la no linealidad del iluminante a lo largo del espectro y eliminar posibles ruidos constantes, se realiza la **normalización** de la reflectancia. Para ello, se captura un patrón que tenga un alto coeficiente de reflexión (máximo) en el rango que se está trabajando, llamado “blanco” y otro patrón con el mínimo coeficiente de reflexión denominado “negro” (Ivorra Martínez, 2015).

La **segmentación** se realiza con el objetivo de separar el fondo de la imagen analizada. El fondo de la imagen no se puede analizar ya que genera espectros con mucho ruido, dificulta el resultado de los modelos multivariantes y ocupa mucho espacio.

Los métodos para eliminar el fondo son, entre otros:

- histogramas basados en PCA,
- puntuaciones (scores) en los diagramas de dispersión de PCA,
- coeficientes específicos de  $\lambda$ ,
- selección manual y,
- K-means minimizando la suma de distancias entre cada observación y el centroide de su grupo o clúster (distancia euclídea).

Con detección de anomalías, nos referimos a la detección de píxeles muertos o líneas causadas por anomalías en el detector y picos irregulares. Por un lado, si aparecen líneas enteras de píxeles muertos, se eliminan por completo, sin embargo, cuando son píxeles dispersos estos se interpolan con el valor de sus vecinos. Por otro lado, los picos irregulares (spikes) son medidas erróneas de una longitud de onda. Estos se deben detectar y eliminar para cada píxel. La detección de estos defectos (píxeles muertos, líneas de píxeles y/o spikes) y su tratamiento puede realizarse por diferentes vías, siendo la detección manual primordial para su tratamiento (Wu y Sun, 2013).

Tras obtener la imagen hiperespectral corregida y seleccionar las regiones de interés (ROI's) tras la segmentación, se realiza el despliegue o **unfolding**. El unfolding consiste en pasar la información almacenada en el hipercubo (3D) a una matriz bidimensional (2D). Este proceso es necesario llevarlo a cabo para poder aplicar posteriormente técnicas de análisis quimiométrico debido a que la mayoría de los algoritmos empleados están diseñados para trabajar en dos dimensiones. Cabe mencionar que, ni los valores de intensidad espectral contenidos dentro del hipercubo, ni el número de elementos del conjunto de datos, cambian. Simplemente se reorganizan los datos en 2D con el propósito de facilitar el análisis. Tal y como muestra la Figura 7, la nueva matriz 2D aporta la reflectancia de cada píxel en cada longitud de onda. El número de filas corresponde con el número de píxeles de la ROI de la muestra y el número de columnas con el número de  $\lambda$  captadas (Dorrepaal et al., 2016).

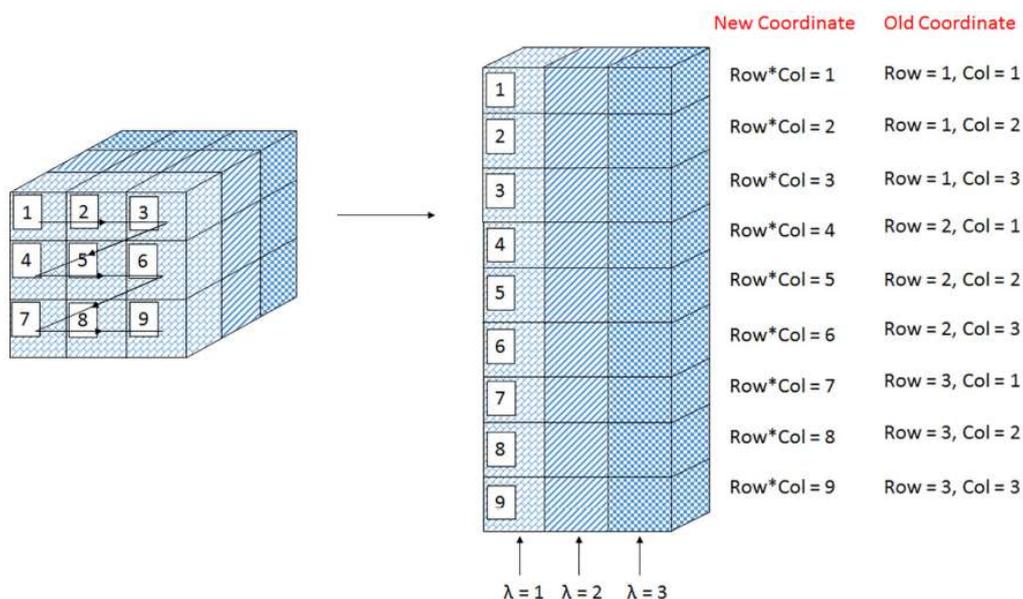


Figura 7. Representación gráfica del despliegue (unfolding) de un hipercubo (Dorrepaal et al., 2016)

Una vez se han generado los espectros, a estos se les realiza un pretratamiento o **preprocesado** para minimizar la dispersión y el ruido espectral. Por un lado, la dispersión se puede minimizar a través de la (1) corrección de la variación normal estándar (*Standard Normal Variate*) –SNV, la (2) corrección multiplicativa de la dispersión (*Multiplicative Scatter Correction*) –MSC o, (3) derivadas. Por otro lado, el ruido espectral se puede minimizar por medio del (1) suavizado (*Smoothing*) o (2) centrado medio (*Mean Centering*). El suavizado de ellos consiste en aplicar una derivada de orden cero a los datos espectrales y el centrado medio en calcular el espectro medio del set de datos para restárselo posteriormente a cada espectro (Dorrepal et al., 2016).

Al finalizar este proceso y todos los tratamientos previos, tal y como se ha citado con anterioridad, la información espectral de la muestra es apta para iniciar el análisis quimiométrico o análisis de datos.

### 2.3.5. Ventajas y desventajas

Siendo conocedores de las características de la técnica HSI en el rango NIR, se procede a detallar las ventajas e inconvenientes de esta técnica. Las imágenes hiperespectrales ofrecen una mayor cantidad de información y una caracterización más fiable de un producto que la de una sola imagen monocromática o una imagen RGB convencional (Elmasry, Kamruzzaman, et al., 2012). Tal y como se ha mencionado, los hipercubos pueden proporcionar información física y química de la muestra analizada. Sin embargo, esta técnica produce un volumen de datos muy grande los cuales sufren problemas de colinealidad y complican el almacenamiento, la gestión y el posterior procesamientos y análisis de las imágenes (Sun, 2010).

En lo que a la industria agroalimentaria respecta, se identifican las siguientes ventajas y desventajas del empleo de la técnica HSI (Tabla 2).

Tabla 2. Ventajas e inconvenientes del empleo de la técnica de imágenes hiperespectrales (Sun, 2010)

Ventajas	Desventajas
Técnica no invasiva ni destructiva de las muestras.	En comparación con los métodos de análisis químicos, las HSI presentan unos límites de detección menores.
Técnica sostenible medioambientalmente al no precisar de químicos para la evaluación de la muestra.	
Análisis rápido y económico de la muestra. En comparación con las técnicas tradicionales, ahorro de: tiempo, mano de obra y costes de los reactivos y de los tratamientos de residuos.	Dificultad de extracción de datos espectrales. Existencia de problemas de multicolinealidad debido a la superposición de bandas.
Mínima preparación de la muestra.	
Capacidad para realizar tanto análisis cualitativos como cuantitativos.	Técnica no aplicable a muestras líquidas y homogéneas al perderse capacidad para relacionar la información espacial.
Delimitación múltiple y simultánea de diferentes componentes dentro de una muestra.	
Proporciona gran cantidad de información útil para la evaluación de la calidad y seguridad alimentaria.	Necesidad de una gran velocidad del hardware para mejorar la adquisición y el análisis de datos.
No se requiere un conocimiento previo de la muestra analizada.	Requiere de calibración precisa y modelos robustos.

## 2.4. Quimiometría

Massart et al. (1988) definió el término quimiometría como la disciplina química que emplea matemática, estadística y otros métodos que utilizan la lógica formal para: (1) diseñar o seleccionar procedimientos de medición y experimentos óptimos y, (2) proporcionar la máxima información química relevante mediante el análisis de datos químicos (Massart, Vandeginste, Deming, Michotte, y Kaufman, 1988). Sin embargo, la primera vez que se introdujo dicho término fue en 1972 por Wold y, casi 30 años después, confirmó junto a Sjöström que la quimiometría se había convertido en una aplicación esencial no sólo en química sino también en otros campos (Wold y Sjöström, 1998).

Como se menciona en el apartado “2.3.5. Ventajas y desventajas”, resulta difícil extraer información de datos espectrales, especialmente en la región NIR, debido a la difícil interpretación de bandas superpuestas. Un espectro NIR es una matriz de datos compleja que resulta de la combinación de diferentes compuestos y propiedades de la muestra estudiada. Por ello, para poder relacionar la información electromagnética (espectro) proporcionada con la información de la composición física y química o los atributos de interés de la muestra analizada, es necesario el uso de la quimiometría (Elmasry, Kamruzzaman, et al., 2012).

La quimiometría se basa en un análisis multivariante a través del cual se descompone de forma eficiente la cantidad masiva de datos almacenados en las imágenes hiperespectrales en información útil y de fácil interpretación (Elmasry, Kamruzzaman, et al., 2012). En definitiva, el análisis quimiométrico es implementado para analizar la información espectral con la finalidad de visualizar la información de interés oculta en las imágenes hiperespectrales. Este proceso es conocido como desarrollo de modelos cualitativos o cuantitativos.

### 2.4.1. Análisis cualitativo

El análisis cualitativo, también conocido como reconocimiento de patrones, se basa en comparar los espectros de muestras desconocidas con los espectros de muestras con características conocidas (Elmasry, Kamruzzaman, et al., 2012). Habitualmente este tipo de análisis se emplean para clasificar las muestras según variables categóricas como, por ejemplo, variedad, origen geográfico o presencia de defectos (Lopez Maestresalas, 2016).

Las técnicas de reconocimiento de patrones (análisis cualitativo) se agrupan en dos clases: (1) reconocimiento de patrones supervisado y (2) no supervisado. Las clases son diferenciadas en base al conocimiento previo de los objetos/muestras que se van a agrupar. Así, los algoritmos no supervisados no requieren ningún conocimiento previo sobre las muestras a agrupar.

A continuación, se recoge una breve descripción de los algoritmos más comunes utilizados en el análisis multivariante de datos espectrales.

#### 2.4.1.1. Técnicas supervisadas

Estas técnicas también son conocidas como análisis discriminantes. Además de la información espectral, se parte del conocimiento previo de información sobre los grupos o categorías existentes en el conjunto de muestras. Estas técnicas generan reglas de clasificación con la finalidad de agrupar las muestras con las que presenten mayor semejanza en términos estadísticos. Es decir, correlacionan la información espectral con la información de la variable del grupo o categoría.

A continuación, se citan las técnicas supervisadas más conocidas empleadas para la realización del análisis cualitativo (Elmasry et al., 2012; Wu y Sun, 2013):

- Análisis de discriminación lineal (LDA)
- Análisis discriminante de mínimos cuadrados parciales (PLS-DA)
- Redes neuronales artificiales (ANN)
- Máquinas de vectores de soporte (SVM)
- k-vecino más cercano (kNN)

#### *2.4.1.2. Técnicas no supervisadas*

A diferencia de las técnicas supervisadas, las no supervisadas no requieren un conocimiento previo sobre los grupos o categorías existentes en el conjunto de muestras. De este modo, en este caso el número de categorías es desconocido.

El objetivo principal de las técnicas no supervisadas empleadas en los análisis cualitativos es identificar las fuentes de agrupación en las muestras disponibles considerando únicamente la información de sus datos espectrales (Lopez Maestresalas, 2016).

Estas técnicas se aplican normalmente en la evaluación preliminar o en las primeras etapas de la exploración de los contenidos de información en el conjunto de datos espectrales de las muestras objeto de estudio. Los análisis de componentes principales (PCA), agrupamiento de K-medias (clúster) y, agrupamiento jerárquico son algunos ejemplos de técnicas no supervisadas empleadas en los análisis cualitativos (Elmasry et al., 2012; Wu y Sun, 2013).

#### **2.4.2. Análisis cuantitativo**

Los modelos de análisis cuantitativos correlacionan la concentración de uno o más compuestos en una muestra, o un atributo físico, químico o biológico de un producto, con la información espectral empleando un modelo matemático. Estos algoritmos se clasifican en lineales y no lineales. A continuación, se recoge una breve descripción de los algoritmos más comunes utilizados en el análisis multivariante de datos espectrales.

##### *2.4.2.1. Métodos de regresión lineal*

Las principales técnicas estadísticas de regresión lineal empleadas para el análisis cuantitativo son: (1) regresión lineal múltiple (MLR), (2) regresión de componentes principales (PCR) y, (3) regresión de mínimos cuadrados parciales (PLS). Estas tres técnicas quimiométricas establecen una relación matemática lineal entre los datos espectrales y el parámetro de la muestra estudiada. Cada una de ellas es descrita a continuación (Elmasry et al., 2012; Lopez Maestresalas, 2016; Wu y Sun, 2013):

- Regresión lineal múltiple (MLR): el algoritmo establece una relación entre el espectro y los atributos deseados de la muestra objeto de estudio en forma de una ecuación lineal. Es decir, establece una relación entre las variables espectrales disponibles y los valores de referencia. Esta técnica falla cuando el número de variables es superior al de muestras y puede verse fácilmente afectada si las variables están correlacionadas.
- Regresión de componentes principales (PCR): se caracteriza por ser un método multivariante de dos pasos. En primer lugar, a través de un análisis de componentes principales (PCA) se descompone la matriz de datos espectrales generando así nuevas variables ortogonales no correlacionadas (PC). En segundo lugar, las PC se emplean en un MLR como variables independientes. El hecho de generar las variables independientes previo a realizar un MLR hace que contengan menos ruido y la técnica no se vea afectada. Sin embargo, no

necesariamente los datos espectrales empleados para calcular las nuevas variables son los más informativos.

- Regresión de mínimos cuadrados parciales (PLS): caracterizada por ser la técnica más robusta y confiable en el desarrollo de un modelo de calibración multivariante. Se basa en predecir un conjunto de variables dependientes, como concentraciones en una muestra, a partir de variables independientes (información espectral). La predicción se consigue extrayendo un conjunto de factores ortogonales llamados variables latentes (LV por sus siglas en inglés) los cuales se caracterizan por tener el mejor poder predictivo.

#### 2.4.2.2. Métodos de regresión no lineal

Las redes neuronales artificiales (ANN) y la regresión vectorial de soporte (SVR) son las principales técnicas de regresión no lineal. Estas técnicas son empleadas cuando no hay una relación lineal entre los datos espectrales y los datos de referencia (atributos objetivo) (Elmasry et al., 2012; Wu y Sun, 2013).

- Redes neuronales artificiales (ANN): caracterizado por simular el comportamiento de las redes de neuronas biológicas para el aprendizaje y el propósito de predicción. La técnica ANN más empleada es la red neuronal de alimentación multicapa.
- Regresión vectorial de soporte (SVR): método basado en el aprendizaje estadístico teórico el cual está basado en dos fases. La primera fase corresponde a la de entrenamiento ya que se les informa con multitud de ejemplos resueltos en forma de pares (problema-solución). La segunda fase está relacionada con la resolución de problemas. En esta última fase la máquina de vectores de soporte (SVR) hace la función de “caja negra” al proporcionar una respuesta (salida) ante un problema dado (entrada).

Cabe destacar que, a pesar de que no haya una relación lineal entre los datos, si se realiza un tratamiento previo de estos, se podrían emplear las técnicas de regresión lineal mencionadas en el apartado “2.4.2.1. Métodos de regresión lineal”.

## 2.5. Aplicaciones HSI en el cultivo de patata.

Los sistemas de imágenes hiperespectrales surgieron en investigaciones relacionadas con la observación de la tierra (Lillesand, Kiefer, y Chipman, 2015). Posteriormente se extendieron a otros campos de la ciencia como la tecnología alimentaria (Wu y Sun, 2013). En relación a dicho campo de trabajo, las imágenes hiperespectrales han sido aplicadas en múltiples trabajos de amplio espectro tales como: predicción de terneza en carne de res, detección de infecciones por mosca de la fruta, predicción de composición en diversos alimentos, detección de defectos y daños en los productos, etc. Así mismo, las imágenes hiperespectrales han sido y están siendo aplicadas en el campo de la agricultura en áreas como: manejo post-cosecha de frutas y hortalizas, evaluación de procesos de deterioro durante el almacenamiento, evaluación del estado de madurez, seguimiento de cambios composicionales y contaminación, entre otras (Wu y Sun, 2013).

En el campo de la agricultura, además, cabe destacar la importancia del sistema de imágenes hiperespectrales en el concepto de agricultura de precisión. La agricultura de precisión se basa en el concepto de aplicar la cantidad correcta de insumos, en el momento adecuado y en el lugar exacto. Es el uso de la tecnología de la información para adecuar el manejo de suelos y cultivos a la variabilidad presente dentro de un lote (García y Flego, 2008). En este caso, la aplicación de nuevas tecnologías como las HSI en la agricultura permite y facilita alcanzar los fundamentos en los que se basa la agricultura de precisión encaminada hacia una sostenibilidad medio ambiental.

Concretamente, en lo que a la agricultura respecta, gracias a los avances tecnológicos se ha conseguido optimizar el uso de diferentes insumos como son los fertilizantes y fitosanitarios. En este caso, a través de las imágenes se puede detectar el estado del cultivo en lo relacionado con condiciones de estrés hídrico o de nutrientes, entre otros. En general, el conjunto de las nuevas tecnologías y sistemas aplicados en la agricultura, como el monitoreo, está facilitando a los agricultores la realización de una buena gestión y manejo preciso de los cultivos con la consiguiente aportación a la mitigación del cambio climático.

A continuación, se citan ciertos estudios en los que se ha aplicado la tecnología de las imágenes hiperespectrales en el cultivo de patata:

- Estimación de la cobertura del suelo y evaluación de la fertilización nitrogenada en papa con registros de cámaras digitales convencionales y modificadas (Casa, Ovando, Bressanini, y Martínez, 2013).
- Detección no destructiva de puntos negros (blackspot) en patatas, con tecnología de imagen hiperespectral VIS-NIR y SWIR (López-Maestresalas et al., 2016).
- Espectroscopía de infrarrojo cercano e imágenes hiperespectrales para el control de calidad no destructiva de patatas (Lopez Maestresalas, 2016).
- Detección de la sarna común en las patatas mediante un sistema de imágenes hiperespectrales infrarrojas (Dacal-Nieto, Formella, Carrión, Vazquez-Fernandez, y Fernández-Delgado, 2011a).
- Detección no destructiva de alteraciones internas (hollow heart) en patatas usando imágenes hiperespectrales (Dacal-Nieto, Formella, Carrión, Vazquez-Fernandez, y Fernández-Delgado, 2011b).
- Detección de la enfermedad del tizón temprano (*Alternaria solani*) en plantas de patata en dos etapas de crecimiento diferentes mediante espectroscopía hiperespectral (Atherton, Watson, Zhang, Qin, y Liu, 2015).
- Clasificación de papa de acuerdo con su aptitud de transformación industrial frito y cocido usando imágenes hiperespectrales (Oliva Lobo, 2020).

### 3. MATERIALES Y MÉTODOS

#### 3.1. Diseño experimental

Con el fin de alcanzar el objetivo del presente estudio se llevó a cabo el procedimiento descrito a continuación.

Se obtuvieron imágenes hiperespectrales de hojas de patata, concretamente, se tomaron imágenes hiperespectrales de 240 hojas de 6 variedades de patata diferentes (20 hojas por tratamiento y variedad). Es decir, por cada variedad, se emplearon 20 hojas sometidas al tratamiento “control” y 20 hojas sometidas al tratamiento “déficit de nitrógeno”.

La adquisición de imágenes se llevó a cabo a través de un sistema hiperespectral de escaneo lineal sensible en el rango NIR, de 900 a 1700 nm. Las imágenes resultantes se procesaron en el software MATLAB® con el objetivo de obtener la información espectral y espacial de cada una de ellas. Para construir el modelo de clasificación se seleccionó aleatoriamente, a través del software, el 80% de las muestras (192 imágenes, 1 imagen por hoja de patata) (grupo de calibración). El 20% restante (48 muestras) fue utilizado para validar externamente el modelo (grupo de validación).

La información obtenida se analizó mediante el método discriminante por mínimos cuadrados parciales (PLS-DA) para los grupos de control frente a deficiencia de nitrógeno. Para ello se utilizó el software PLS\_Toolbox versión 8.9, el cual trabaja en el entorno MATLAB®. El análisis PLS-DA fue llevado a cabo con el fin de generar un modelo de clasificación de hojas de plantas de patata sometidas a diferentes tratamientos de fertilización nitrogenada, sin diferencias visibles, en base a sus diferencias espectrales.

### 3.2. Material vegetal

El material vegetal empleado para el presente estudio fue cultivado, seleccionado y recogido de forma manual en una finca experimental en Arkaute (Álava) perteneciente al Instituto Vasco de Investigación y Desarrollo Agrario (NEIKER).

Pese a que para el estudio se emplearon plantas de 6 variedades diferentes, las plantas pertenecían a un estudio varietal más amplio (150 variedades) con el objetivo de identificar qué variedades se adaptaban mejor a condiciones de cultivo más restrictivas desde el punto de vista del abonado nitrogenado y del riego. Para ello se realizaron 3 ensayos de plantas de patata, cada uno de ellos sometido a un tratamiento diferente: (1) “control”, (2) “estrés hídrico”, (3) “déficit de nitrógeno”. Cada ensayo comprendía 150 variedades (las mismas para los 3 ensayos) y 2 repeticiones de 5 plantas cada repetición.

De estas 150 variedades que constituían el estudio varietal, las 6 variedades seleccionadas para el presente estudio fueron: (1) Alibaba, (2) Baraka, (3) Laudine, (4) Mayka, (5) Romani y, (6) Ventana. Estas 6 variedades fueron seleccionadas en base al estado fenotípico de las mismas. De esta forma, se escogieron 2 variedades catalogadas con buen aspecto, 2 variedades con mal aspecto y 2 variedades intermedias, fenotípicamente hablando. Así mismo, en base al objetivo de este estudio, las plantas analizadas fueron las sometidas a los tratamientos “control” y “déficit de nitrógeno”. Sin embargo, se llevó a cabo, de forma paralela, otra investigación la cual estudió las plantas sometidas a los tratamientos “control” y “estrés hídrico”. Todos los ensayos fueron sembrados el 15 de mayo de 2019.

En lo que a la cantidad de agua aportada respecta, no existieron diferencias entre el tratamiento “control” y “déficit de nitrógeno”. En ambos ensayos se realizaron 10 riegos a lo largo de todo el ciclo de cultivo. Los riegos se espaciaron una media de 6 días y aportaron una media de 13,7 l/m<sup>2</sup> cada uno. El primer y último riego se realizó el 24 de junio y 23 de agosto de 2019, respectivamente. Además, ambos tratamientos se vieron beneficiados por la precipitación natural. La precipitación recogida en la estación meteorológica de Arkaute desde el inicio del ciclo de cultivo hasta el 15 de julio, momento de recolección de las hojas fue de 58 l/m<sup>2</sup>.

La diferencia entre ambos tratamientos radica en el abonado nitrogenado. Por un lado, las plantas sometidas al tratamiento “control” fueron abonadas en 2 ocasiones a lo largo del ciclo del cultivo. En las 2 ocasiones de abonado se aplicaron 90 unidades de nitrógeno con un abono complejo NPK 9-18-27. Sin embargo, la primera aplicación fue un abonado de fondo y la segunda, ya con el cultivo en crecimiento, fue un abonado de cobertera. Por otro lado, las plantas sometidas al tratamiento “déficit de nitrógeno” fueron abonadas una única vez. En este caso se aportó un abonado de fondo al inicio del ciclo del cultivo con la mitad de la dosis, es decir, se aplicaron 45 unidades de N. Es decir, se redujo un 75% el aporte de Nitrógeno.

El 15 de julio del mismo año se recolectaron las 240 hojas diferentes (20 hojas por tratamiento y variedad) y se enviaron al Laboratorio de Ingeniería de la Posrecolección de Productos Agroalimentarios de la Escuela Técnica Superior de Ingeniería Agronómica y Biociencias de la Universidad Pública de Navarra para el análisis de imagen, el cual se realizó el mismo día.

### 3.2.1. Identificación de las muestras

Para garantizar un buen control y trazabilidad de los datos fue esencial realizar una correcta identificación de las muestras de material vegetal recogidas en campo. En este estudio la nomenclatura seguida para la identificación de las muestras fue la siguiente:

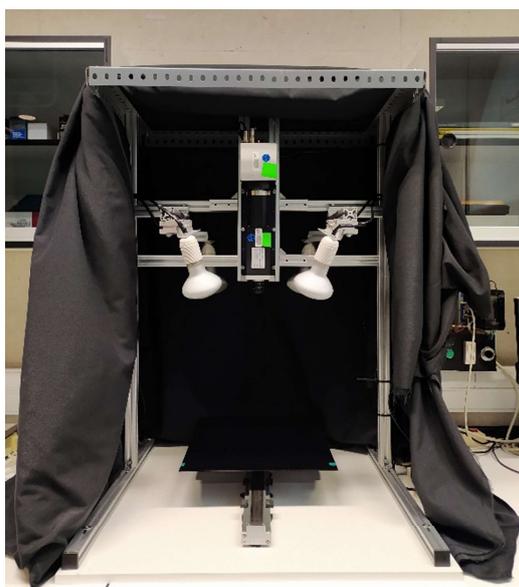
01 024 06

a b c

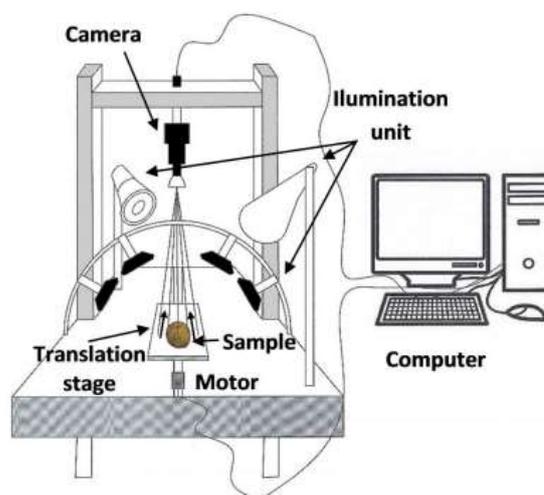
- a. *Tratamiento* → 01: “control”; 03: “déficit de nitrógeno”.
- b. *Variedad* → 010: Alibaba; 024: Baraka; 075: Laudine; 091: Mayka; 129: Romani; 147: Ventana.
- c. *Nº de hoja* → hojas de la 1 a la 20 (01-20) por variedad y tratamiento.

### 3.3. Adquisición de imágenes hiperespectrales (HSI)

Las imágenes hiperespectrales de las hojas se adquirieron en el Laboratorio de Ingeniería de la Posrecolección de Productos Agroalimentarios de la Escuela Técnica Superior de Ingeniería Agronómica y Biociencias de la Universidad Pública de Navarra. Este laboratorio cuenta con un equipo de análisis de imagen (Figura 8) mediante el cual se escanearon las hojas seleccionadas y recolectadas de las diferentes plantas de patata. Este equipo está compuesto por un sistema hiperespectral de escaneo lineal sensible en el rango NIR 900-1700 nm, una plataforma móvil, una fuente de iluminación y un ordenador.



(a)



(b)

Figura 8. Equipo de adquisición de imágenes hiperespectrales empleado: equipo real (a); diagrama esquemático (b) (Lopez Maestresalas, 2016)

El sistema hiperespectral empleado incluye una cámara Xeva 1.7-320-100Hz (Xenics, Leuven, Bélgica) de escaneo lineal, sensible en el rango NIR de 900-1700 nm. La cámara incorpora un detector InGaAs de resolución 320 x 256 píxeles y conexión USB. La cámara está acoplada a un espectrógrafo ImSpector N17E (Specim, Spectral Imaging Ltd., Oulu, Finlandia) con una abertura (slit) de 30 µm, y a una lente OPT-000034 (SWIR, 16, mm. f/1.4, Xenics, Leuven, Bélgica) con una distancia focal de 16 mm (Figura 9).

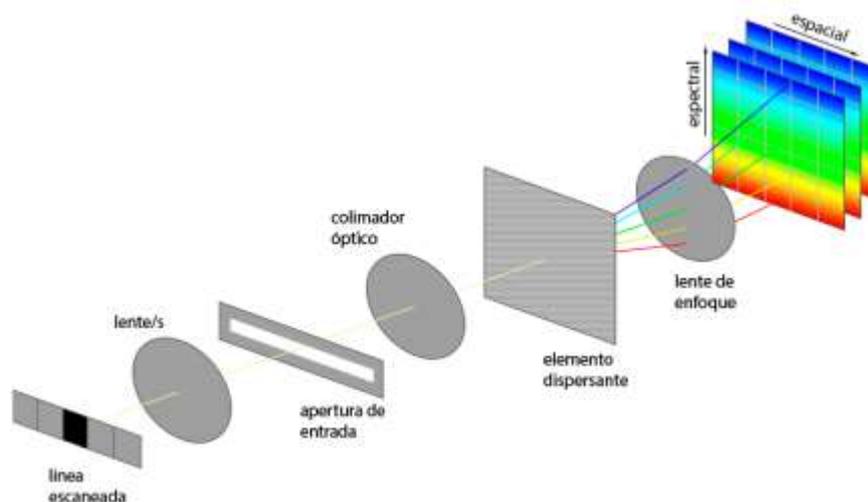


Figura 9. Esquema del funcionamiento del espectrógrafo en el sistema de adquisición de imágenes hiperespectrales (Municio, 2011)

Para mover la muestra a velocidad constante a través de la zona de escaneo, es decir, del campo de visión de la cámara, se empleó una plataforma móvil lineal (LEFS25, SMC Corporation, Tokio, Japan) movida por un actuador eléctrico con un motor de pasos (Servo/24 VDC), y controlado a través de un ordenador por medio del software ACT Controller. Para el escaneo de las muestras se fijó una velocidad de la plataforma de aproximadamente 7 mm/s y una distancia de separación entre la lente y la superficie de la muestra de 40 cm. Además, se ajustó una distancia de enfoque de 25 cm y la máxima apertura del diafragma ( $f/1.4$ ) en el lente. Las imágenes se adquirieron en el modo de calibración TrueNUC-HG de la cámara, a la máxima velocidad de escaneo (100 Hz), y con un tiempo de integración de 2.000  $\mu\text{s}$ . De este modo, las dimensiones de las imágenes resultantes fueron las siguientes: 320 columnas (0,75 mm/pixel), 256  $\lambda$  (cada 3,125 nm) y un número variable de filas (en función del tamaño de la hoja, filas escaneadas).

La fuente de iluminación empleada estaba constituida por 4 lámparas halógenas de 28W (Lexman). Estas emiten radiación en el espectro infrarrojo. Con el fin de conseguir una iluminación homogénea en el campo visual de la cámara y evitar un exceso de iluminación que pudiera dar lugar a imágenes saturadas, las lámparas fueron colocadas de tal forma que cada una de ellas enfocaba una esquina de la placa portamuestras. Además, con el objetivo de evitar interferencias por radiaciones externas de la luz ambiental y obtener datos erróneos, se cubrió todo el sistema con una tela negra opaca durante el escaneo de las hojas.

Para controlar el sistema de adquisición hiperespectral y guardar las imágenes generadas se empleó un ordenador equipado con el software Xeneth 2.5. Este software permite establecer los parámetros de ensayo de la cámara y controlar su funcionamiento. El ordenador contaba además con el software MATLAB® a través del cual se realizó el análisis y procesado de las imágenes hiperespectrales.

Por último, se realizó la normalización de la reflectancia capturando patrones de alto y bajo coeficiente de reflexión para corregir la reflectancia relativa de las imágenes con el máximo y mínimo nivel de reflectancia que es capaz de captar el detector (cámara). A continuación, se detallan los 2 niveles de reflectancia (máximo y mínimo):

- Blanco: máximo nivel de reflectancia que es capaz de detectar la cámara. Para adquirir las imágenes que definen dicho nivel de reflectancia se colocó un estándar de referencia blanco *Spectralon* (Standard 99%, Labsphere) en la placa portamuestras con el sistema de iluminación encendido.
- Negro: mínimo nivel de reflectancia que es capaz de detectar la cámara. Para adquirir las imágenes que definen dicho nivel de reflectancia deben darse las condiciones de mínima radiación. Estas se consiguieron colocando la tapa en la lente y apagando el sistema de iluminación.

### 3.4. Procesado de imágenes

Tras la adquisición de las imágenes hiperespectrales, se procesaron empleando MATLAB R2020b (The Mathworks, Natick, MA) mediante código propio desarrollado específicamente para este fin. El procesado es llevado a cabo con la finalidad de extraer la información útil de las muestras en el formato adecuado para, posteriormente, analizar la información. Para el procesado, en primer lugar, se formaron los hipercubos (1 hipercubo/imagen) y se normalizaron. Seguidamente, se segmentaron las imágenes para separar las regiones de interés, en este estudio la hoja completa, del fondo. A continuación, se reordenó la información almacenada en el hipercubo (3D) en una matriz bidimensional (2D) almacenando la reflectancia de cada pixel en cada longitud de onda (*unfolding*). Por último, se calculó el espectro medio de cada hoja del grupo de calibración para posteriormente llevar a cabo el análisis discriminante (véase apartado “4.5. Análisis quimiométrico: análisis multivariante”).

#### 3.4.1. Formación del hipercubo

A partir de las imágenes hiperespectrales obtenidas de cada muestra (hoja de patata) se construyó el hipercubo. Es decir, en este estudio se formaron 240 hipercubos en total.

El hipercubo de cada muestra se formó apilando las imágenes adquiridas por la cámara. De esta forma, a partir de un número limitado de imágenes (tantas como filas se escanearon por muestra), se obtuvo un hipercubo por muestra cuyas dimensiones fueron 320 columnas x 256  $\lambda$  x n° de filas como imágenes fueron escaneadas. Así, la altura del hipercubo siempre es función del tamaño de la muestra (Figura 2).

#### 3.4.2. Normalización

Tras la obtención del hipercubo de la muestra, se llevó a cabo la normalización de los valores de reflectancia obtenidos. Este paso se realizó para facilitar el estudio debido a que los valores de reflectancia obtenidos en las imágenes variaban entre 0 y  $2^{16}$  (65536) niveles de grises. Así, tras la normalización, los valores de reflectancia de las imágenes pasaron a estar comprendidos entre 0 y 1.

Para escalar todos los valores de reflectancia entre 0 y 1 se emplearon como referencia las imágenes correspondientes al máximo (blanco) y mínimo (negro) nivel de reflectancia que era capaz de captar el receptor (véase el apartado “3.3. Adquisición de imágenes hiperespectrales (HSI)”). La normalización se llevó a cabo según la Ecuación 1:

$$R_r = \frac{R_m - R_n}{R_b - R_n}; \quad (1)$$

donde,

$R_r$ : reflexión relativa

$R_m$ : reflexión de la muestra

$R_n$ : reflectancia mínima o negro

$R_b$ : reflectancia máxima o blanco

### 3.4.3. Segmentación

Una vez se obtuvieron los hipercubos normalizados, se dividió la información en 2 grupos. Se empleó el 80% de la información (192 muestras) para crear un grupo de calibración y el 20% restante para crear un grupo de validación externa (48 muestras). La información del grupo de validación externa fue empleada para comprobar la eficacia del modelo de calibración para clasificar las muestras objeto de estudio. La división de la información en los 2 grupos (calibración y validación) se realizó de forma aleatoria mediante un algoritmo del software MATLAB®.

Tras la división de la información, se procedió a seleccionar las regiones de interés, es decir, aquellas regiones de cada hipercubo con información espectral útil desde el punto de vista espacial. Este proceso es llevado a cabo debido a que no todos los píxeles de la imagen proporcionan información espectral útil.

A través de la segmentación de la imagen se separó la información de interés de la muestra de aquella sin utilidad desde el punto de vista espacial. En este caso, se separó la información aportada por el fondo (plataforma) de la correspondiente con la hoja de patata.

Para obtener las regiones de interés en este estudio se aplicó únicamente una máscara o imagen binaria la cual se combinó con el hipercubo normalizado. En la máscara se asignó un valor de 0 a los píxeles del fondo (plataforma) y un valor de 1 a los píxeles correspondientes con la muestra, siendo estos últimos las regiones de interés (ROI's). Como resultado se obtuvo una imagen de la hoja de patata aislada la cual fue utilizada entonces como región de interés en los siguientes pasos del procesado de imágenes.

### 3.4.4. Unfolding

Por último, se reordenó la información almacenada en cada hipercubo normalizado (3D) en una matriz bidimensional (2D) la cual almacenó los espectros de reflectancia de cada píxel de la hoja de patata en cuestión. Este proceso se realizó combinando la máscara con el hipercubo normalizado resultando así una matriz 2D para cada muestra. Al combinar cada máscara con su hipercubo normalizado correspondiente los píxeles del fondo tomaron valores de 0 mientras que los de la hoja mantuvieron su valor de reflectancia.

Una vez realizado el *unfolding* de cada muestra, se obtuvieron 240 matrices constituidas por tantas filas como píxeles componían cada hoja y, en lo que a las columnas se refiere, tantas como longitudes de onda captadas (256). Sin embargo, debido al gran ruido instrumental que apareció en las primeras longitudes de onda de los espectros, concretamente entre los 900-1000 nm, se eliminaron para cada muestra las 33 primeras  $\lambda$ . De esta forma, se emplearon los valores de reflectancia obtenidos comprendidos entre 1000 y 1700 nm. Por ello, se redujeron de 256 a 223 los valores de reflectancia de cada espectro obteniendo finalmente matrices con tantas filas como píxeles componían cada hoja y 223 columnas. La diferencia entre el espectro con ruido instrumental y sin ruido instrumental se muestra en la Figura 10 y Figura 11, respectivamente.

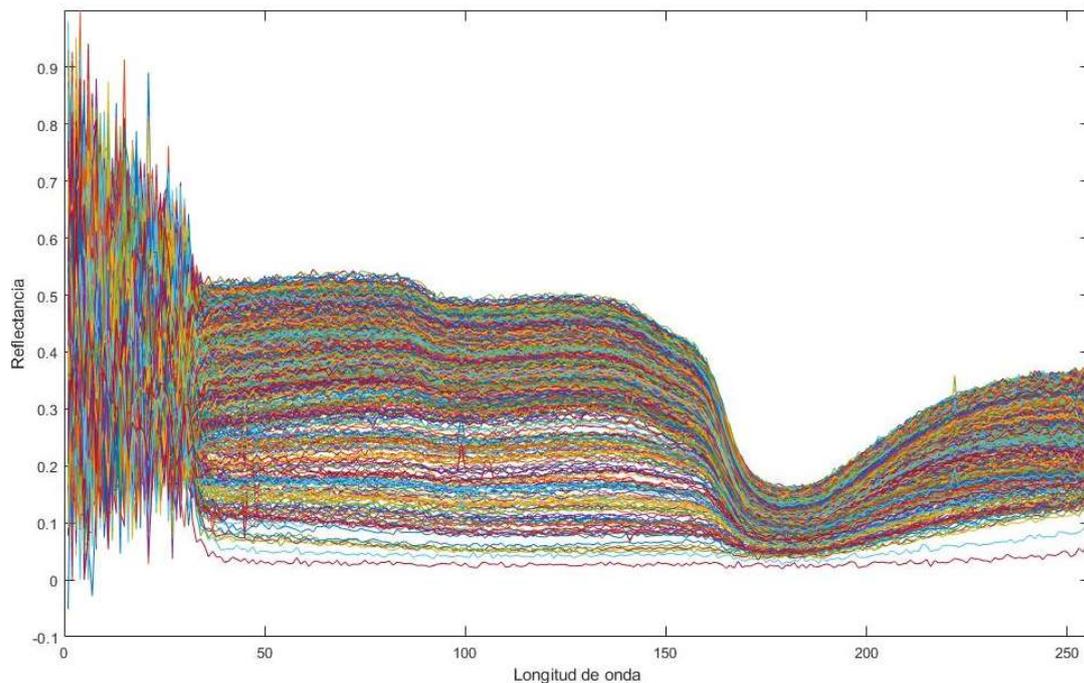


Figura 10. Espectro con ruido instrumental

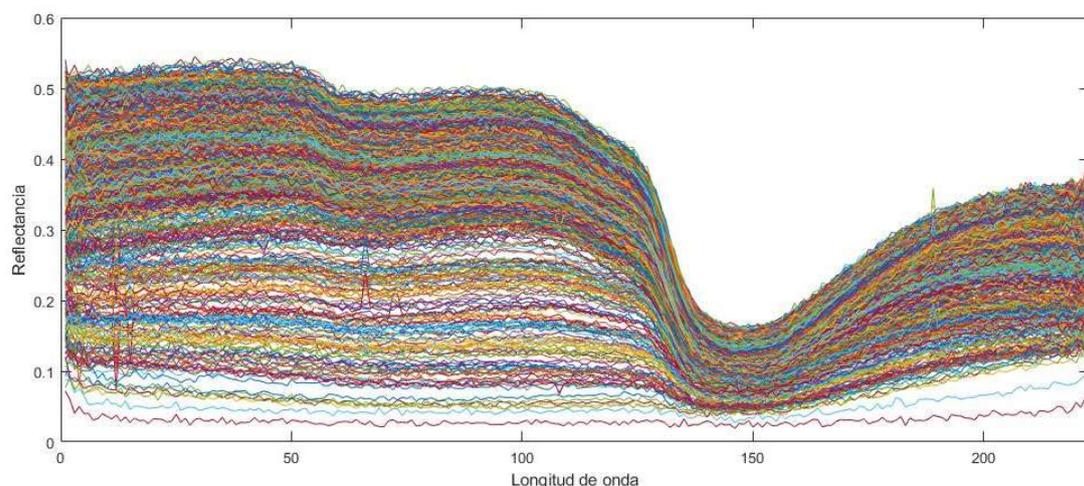


Figura 11. Espectro sin ruido instrumental tras su eliminación

Previamente a la realización del análisis quimiométrico, los datos de las muestras constituyentes del grupo de calibración y los del grupo de validación fueron agrupados en matrices diferentes. Tanto la matriz de datos del grupo de calibración como la del grupo de validación se realizaron mediante el software MATLAB®.

Para el modelo de calibración, se emplearon únicamente los datos de los espectros medios de cada muestra. Por ello, previo al análisis quimiométrico, se agrupó la información de los espectros medios de cada muestra (192) en una matriz formada por 192 filas y 223 columnas.

En el caso del modelo de validación, se emplearon los datos de todos los espectros de cada muestra con el fin de poder crear una imagen química para obtener un mapa representativo de cada una de ellas. De esta forma, la matriz resultante, la cual recogía toda la información espectral, estaba constituida por 115833 filas y 223 columnas. El número de filas resultó de concatenar los 48 *unfolded* de las muestras constituyentes del grupo de validación.

### **3.5. Análisis quimiométrico: análisis multivariante**

Una vez agrupados los datos en matrices y, con el fin de alcanzar el objetivo de clasificación de las hojas de patatas en cada clase (control/déficit de nitrógeno) a través de sus espectros NIR, se empleó la herramienta PLS\_Toolbox. PLS\_Toolbox es un software de análisis quimiométricos multivariante que trabaja bajo el entorno del software MATLAB®.

Para llevar a cabo el análisis quimiométrico de los datos se siguieron los siguientes pasos: (1) selección del método de análisis, (2) selección de cada clase, (3) selección de los pretratamientos de datos, (4) validación cruzada, (5) construcción del modelo y, (6) validación externa. El procedimiento se plasma en la Figura 12 como un diagrama de flujo.

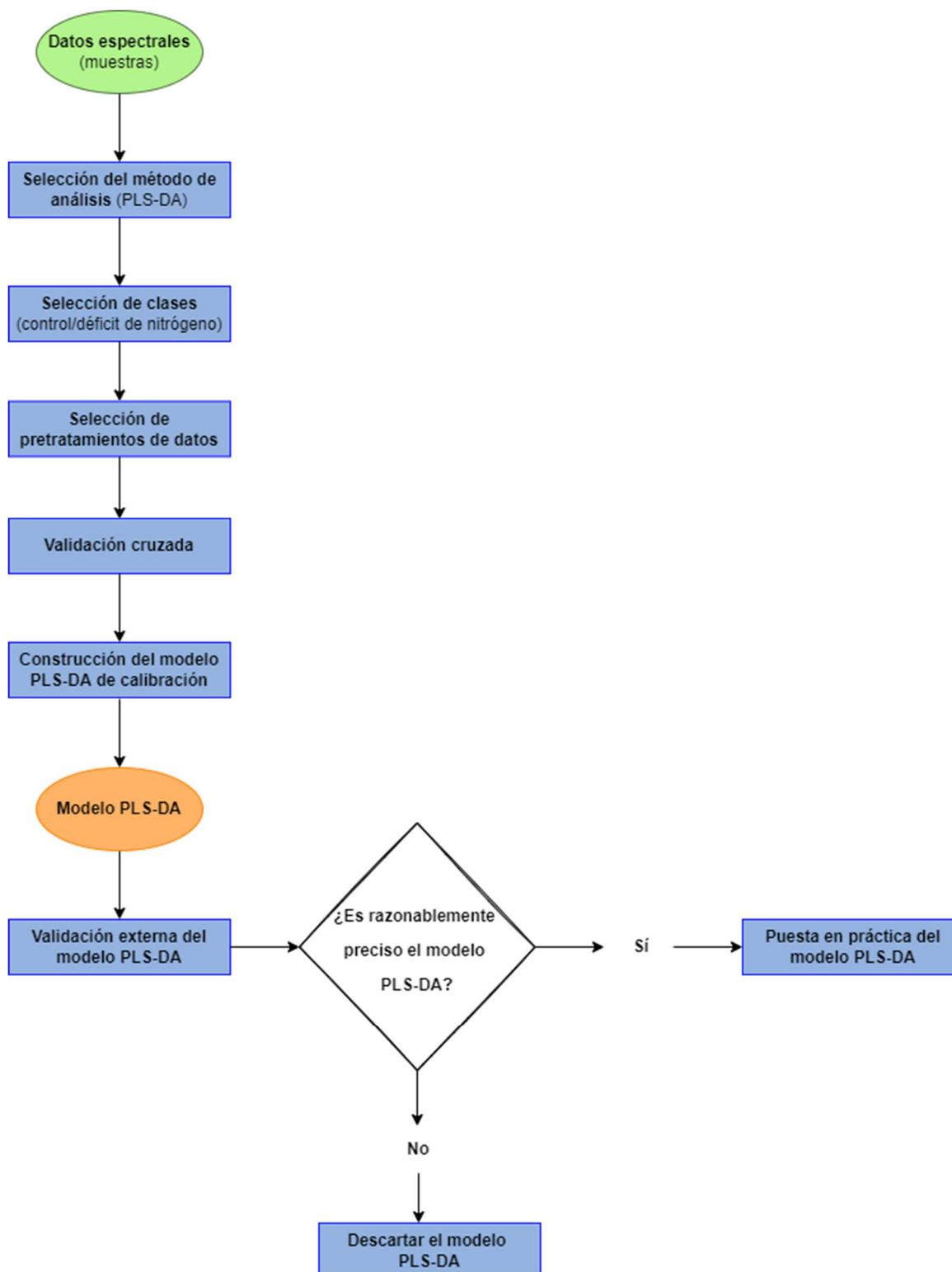


Figura 12. Diagrama de flujo del procedimiento del análisis quimiométrico de los datos espectrales

El método escogido para realizar el análisis quimiométrico de los datos en PLS\_Toolbox fue el método de clasificación supervisado PLS-DA (*Partial Least Square Discriminant Analysis*). El PLS-DA se trata de una técnica (método cualitativo de análisis estadístico) de reconocimiento de patrones donde la pertenencia a una clase se predice a partir de los espectros de las muestras por medio de la regresión lineal por mínimos cuadrados parciales (PLS) (Ceballos et al., 2019).

De este modo, el modelo estableció una relación entre las variables predictivas X (valores de reflectancia en las longitudes de onda medias) y la variable dependiente Y (clases en los datos).

La principal ventaja del PLS-DA es que las fuentes de variabilidad de los datos se modelan mediante variables latentes (combinaciones lineales de las variables originales), y las puntuaciones PLS asociadas se calculan y se representan por pares, lo que permite una evaluación visual de la separación de las clases (de Almeida et al., 2013). El modelo también calcula la probabilidad de que una muestra pertenezca a la clase modelada.

Tras haber seleccionado el método de análisis, se determinaron los pretratamientos que iban a ser aplicados a los datos espectrales. La aplicación de pretratamientos fue realizada con la finalidad de que los modelos de calibración creados fueran capaces de identificar, con mayor facilidad, la clase a la que pertenecía cada hoja de patata en base a la información espectral analizada. Los pretratamientos que se aplicaron a los datos se citan a continuación:

- Suavizado:
  - *SM (Smoothing Savitzky-Golay)*
- Normalización ponderada:
  - *SNV (Standard Normal Variate)*
  - *MSC (Multiplicative Scatter Correction)*
- Primera y segunda derivada de los datos:
  - *Derivate Savitzky-Golay*
- Proceso de eliminación de los valores medios:
  - *MC (Mean Centering)*

Los pretratamientos fueron aplicados tanto de forma individual como combinada a los datos del grupo de calibración recogidos en la matriz de 192 filas y 223 columnas. Tal y como se muestra en la Tabla 3, se crearon 13 modelos diferentes en base a los pretratamientos de datos aplicados.

Tabla 3. Pretratamientos de los modelos de calibración desarrollados (MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada)

Modelo	Pretratamientos
1	NONE
2	MC
3	SNV+MC
4	MSC+MC
5	SM+MC
6	SNV+SM+MC
7	MSC+SM+MC
8	SNV+1ªD+MC
9	SNV+2ªD+MC
10	MSC+1ªD+MC
11	MSC+2ªD+MC
12	SM+MSC+MC
13	SM+SNV+MC

A continuación de la selección de los pretratamientos, se llevó a cabo, en cada modelo, un proceso de validación cruzada. La validación cruzada divide en “n” partes el conjunto de datos y, a continuación, extrae 1 parte como conjunto de prueba para validar el modelo entrenado con las “n-1” partes restantes (FH Joanneum, 2005). Este proceso se repite de forma sucesiva hasta que se han utilizado como conjunto de prueba todas las partes.

Esta técnica, además de medir la precisión de cada modelo, fue empleada con el fin de evaluar la variabilidad del conjunto de datos del grupo de calibración y la confiabilidad de cualquier modelo entrenado con ellos. Además, la validación cruzada permitió tener una idea sobre la representatividad del conjunto de datos y el grado de vulnerabilidad del modelo a variaciones en estos.

Una vez realizado el proceso de validación cruzada, el número de variables latentes que se consideraron óptimas para el desarrollo de cada modelo fue establecido en base al modelo que minimizaba tanto el error cuadrático medio de calibración como el error cuadrático medio de validación cruzada. De esta forma, cada modelo se eligió teniendo en cuenta la mínima distancia entre errores y el mayor porcentaje de varianza explicada. En la Figura 13 se muestra un ejemplo del gráfico empleado para seleccionar el número de variables latentes óptimo para el desarrollo de un modelo de calibración. En este ejemplo particular se seleccionarían 5 LV.

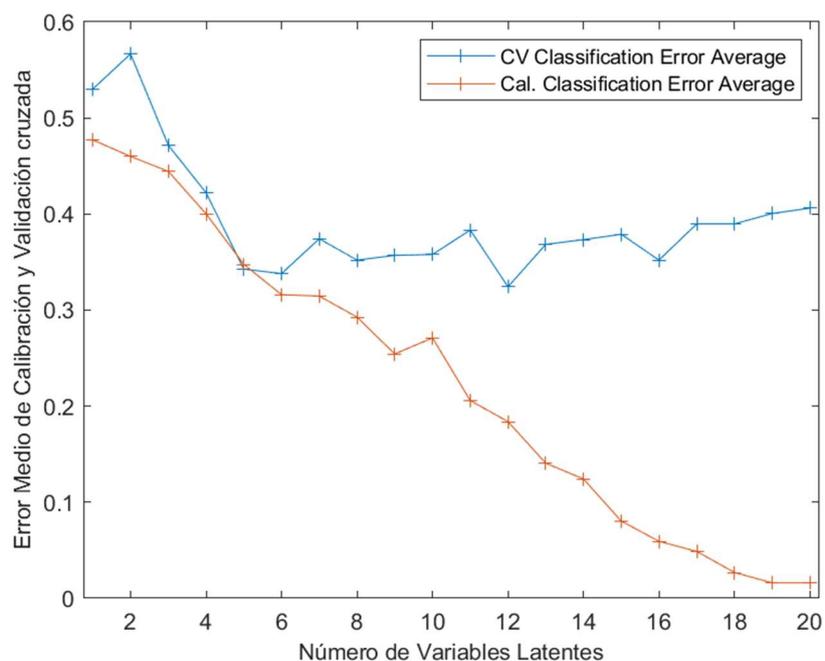


Figura 13. Representación del error cuadrático medio de calibración y validación cruzada en función de las variables latentes (LV)

Tras la selección del número de variables latentes óptimas, se evaluó la efectividad de los modelos de calibración en base a la información obtenida de las correspondientes matrices de confusión (Tabla 4). La matriz de confusión trata de explicar la cantidad de predicciones que se han clasificado correcta e incorrectamente. La matriz de confusión recoge en su diagonal principal los valores que han sido predichos correctamente. En este caso, refleja la cantidad de muestras que han sido clasificadas correctamente en la clase a la que pertenecen. Según la clase, son denominados “Verdaderos Positivos” o “Verdaderos Negativos”. Así, los valores que se encuentran fuera de la diagonal principal son denominados “Falsos Positivos” o “Falsos Negativos” en función de la clase a la que pertenecen. A continuación, se describen brevemente cada uno de estos valores:

- Verdadero Positivo (VP): Cantidad de muestras de la clase “A” clasificadas correctamente en la clase “A”. En este caso, los verdaderos positivos son aquellas muestras de la clase “control” que han sido correctamente clasificadas como pertenecientes a la clase “control”.
- Falso Negativo (FN): Cantidad de muestras de la clase “A” clasificadas incorrectamente como clase “B”. Es decir, aquellas muestras que, aun perteneciendo a la clase “control”, han sido clasificadas como clase “déficit de nitrógeno”.
- Verdadero Negativo (VN): Cantidad de muestras de la clase “B” clasificadas correctamente en la clase “B”. En este caso, los verdaderos negativos son aquellas muestras de la clase “déficit de nitrógeno” que han sido correctamente clasificadas en su clase.
- Falso Positivo (FP): Cantidad de muestras de la clase “B” que han sido clasificadas incorrectamente como clase “A”. Es decir, aquellas muestras que, aun perteneciendo a la clase “déficit de nitrógeno” han sido clasificadas como clase “control”.

Tabla 4. Matriz de confusión binaria

		VALORES REALES	
		Control	Déficit de nitrógeno
VALORES PREDICHOS	Control	Verdaderos Positivos	Falsos Positivos
	Déficit de nitrógeno	Falsos Negativos	Verdaderos Negativo

A partir de los valores anteriormente descritos se define la sensibilidad y la especificidad de los modelos de calibración desarrollados. Ambos parámetros nos indican la capacidad del modelo para discriminar las muestras “control”, de las “déficit de nitrógeno”.

- Sensibilidad: Conocida también como Tasa de Verdaderos Positivos (*True Positive Rate*). Es la proporción de verdaderos positivos (muestras de la clase “control”) que fueron correctamente identificadas por el algoritmo. Se obtiene a través de la Ecuación 2.

$$\text{Sensibilidad} = \frac{VP}{VP+FN} ; \quad (2)$$

- Especificidad: Conocida también como la Tasa de Verdaderos Negativos (*True Negative Rate*). En este caso, se trata de las muestras de la clase “déficit de nitrógeno” que el algoritmo ha clasificado correctamente. Expresa cuan bien puede detectar esa clase el modelo. Se obtiene a través de la Ecuación 3.

$$\text{Especificidad} = \frac{VN}{VN+FP} ; \quad (3)$$

Para ambos parámetros, cuanto menor es el valor obtenido, mayores dificultades tiene el modelo para diferenciar las muestras de cada clase, es decir, discriminar las muestras “control” de las “déficit de nitrógeno”. Además, en base a la definición de cada parámetro, cabe mencionar que, para cada modelo, el valor obtenido de sensibilidad de una clase es igual al valor de especificidad de la otra clase. En este caso, para un modelo dado, el valor de sensibilidad calculado para la clase “control” es igual al valor de especificidad calculado para la clase “déficit de nitrógeno”. De tal forma, el error de clase, descrito a continuación, toma siempre el mismo valor independientemente de la clase para la que se esté calculando. Estas relaciones se pueden comprobar en el apartado “3. Resultados de sensibilidad, especificidad y error de los modelos PLS-DA desarrollados” del Anexo.

Mediante la sensibilidad y la especificidad se calcula el error de clase para cada modelo. Este parámetro es el promedio de las tasas de Falsos Negativos y Falsos Positivos para cada clase. Se obtiene a través de la Ecuación 4.

$$\text{Error de clase} = 1 - \frac{\text{Sensibilidad} + \text{Especificidad}}{2};$$

Además de la sensibilidad y la especificidad de cada modelo, a partir de la matriz de confusión se puede hallar también el porcentaje de muestras correctamente clasificadas (CC). Se obtiene a través de la Ecuación 5.

$$CC = \frac{VP + VN}{\text{Número de muestras clasificadas}}; \quad (5)$$

Finalmente, una vez se obtuvieron los diferentes y definitivos modelos de calibración, el último paso fue realizar su validación externa. Para ello se empleó la matriz elaborada mediante el software MATLAB® (115833 filas x 223 columnas) la cual recogía toda la información espectral de las 48 muestras del grupo de validación (muestras no consideradas en el desarrollo del modelo de calibración). Debido a que el conjunto de datos presentaba mucho ruido, se le aplicaron 2 pretratamientos previo a realizar la validación externa: (1) suavizado (*Smoothing Savitzky-Golay*) y, (2) normalización ponderada con SNV (*Standard Normal Variate*).

## 4. RESULTADOS Y DISCUSIÓN

En el presente estudio se desarrollaron 13 modelos diferentes mediante el análisis PLS-DA de los datos constituyentes del grupo de calibración (192 muestras). Primeramente, los modelos fueron validados internamente a través del proceso de validación cruzada. Posteriormente, se validaron externamente con los datos de las 48 muestras destinadas a la validación externa.

Los resultados obtenidos de la aplicación de las imágenes hiperespectrales para la clasificación de patatas según los efectos, en hojas, de la fertilización nitrogenada se exponen en el presente apartado. En primer lugar, se presentan los resultados de los modelos PLS-DA de calibración generados para la clasificación de las hojas de patatas. En segundo lugar, se contrastan los resultados obtenidos con los del modelo de referencia. En tercer lugar y, por último, se recogen los resultados relacionados con la validación externa de estos modelos.

### 4.1. Modelos PLS-DA para clasificar hojas de patata según los efectos de la fertilización

La diferencia entre los modelos de calibración desarrollados radicó en los diferentes pretratamientos aplicados a los datos espectrales. Los diferentes pretratamientos aplicados en cada modelo quedan descritos en el apartado “3.5. Análisis quimiométrico: análisis multivariante” y en la Tabla 3 del mismo inciso.

Con el objetivo de poder evaluar los modelos de calibración y, seleccionar aquellos con mejores resultados de clasificación, se registraron para cada modelo, tanto las características de estos como los resultados obtenidos. Todo ello se muestra agrupado en la Tabla 5 la cual recoge: (1) número de modelo, (2) número de muestras empleadas para el desarrollo del modelo, (3) número de *outliers* (muestras eliminadas), (4) pretratamientos aplicados a los datos espectrales, (5) varianza explicada por el modelo, (6) número de variables latentes empleadas (LV), (7) porcentaje de muestras correctamente clasificadas (CC) en la validación cruzada (CV) por cada clase y, grupalmente.

Cabe mencionar que, el número de muestras eliminadas y, por ende, no consideradas para el desarrollo del modelo de calibración en cuestión, supuso una pérdida de datos inferior al 10% del total.

Tabla 5. Características de los modelos de calibración desarrollados. Resaltados los modelos obtenidos más precisos (CC: Correctamente Clasificadas; CV: Validación Cruzada; LV: Variables Latentes; NONE: crudo; MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada)

Modelo	Nº muestras	Outliers	Pretratamientos	Varianza	LV	% CC (CV)		
						Control	Déficit nitrógeno	Grupal
1	185	7	NONE	100	5	58,89	73,68	66,49
2	189	3	MC	99,96	7	56,84	69,15	62,96
3	189	3	SNV+MC	99,15	4	51,04	77,42	64,02
4	182	10	MSC+MC	99,27	5	57,95	71,28	64,84
5	190	2	SM+MC	99,98	6	52,63	78,95	65,79
6	187	5	SNV+SM+MC	99,55	5	51,06	76,34	63,64
7	188	4	MSC+SM+MC	99,74	6	56,38	70,21	63,30
8	189	3	SNV+1ªD+MC	99,64	5	55,79	79,79	67,72
9	187	5	SNV+2ªD+MC	98,99	5	62,77	73,12	67,91
10	185	7	MSC+1ª D+MC	99,70	6	63,04	67,74	65,41
11	188	4	MSC+2ªD+MC	98,61	5	58,51	71,28	64,89
12	181	11	SM+MSC+MC	99,85	7	62,07	72,34	67,40
13	185	7	SM+SNV+MC	99,92	9	67,39	69,89	68,65

Tal y como se ha mencionado anteriormente, la diferencia entre los modelos de calibración que se generaron radicó en los pretratamientos aplicados a los datos espectrales. Es importante resaltar la influencia de los pretratamientos en los resultados que se obtuvieron con cada modelo. No solo variaron los % CC según los pretratamientos aplicados, sino que también variaron en función del orden de aplicación de estos (Tabla 5). En el presente documento se puede detectar esta influencia comparando los resultados obtenidos de %CC de los modelos número 6 y 13 así como los de los modelos número 7 y 12. Los modelos número 6 y 13 se desarrollaron tras aplicar a los datos objeto de estudio los pretratamientos *Standard Normal Variate*, *Smoothing* y *Mean Centering*. Sin embargo, debido a la influencia del orden de aplicación de estos, el modelo número 13 fue capaz de clasificar correctamente aproximadamente un 5% más de muestras que el modelo número 6. Este mismo hecho ocurrió en los modelos 7 y 12 ante la aplicación de *Multiplicative Scatter Correction*, *Smoothing* y *Mean Centering* en diferente orden. El modelo 12 fue capaz de clasificar correctamente aproximadamente un 4% más de muestras que el modelo 7. De tal forma se corrobora la influencia tanto de los pretratamientos aplicados como del orden de aplicación.

La influencia, en los resultados obtenidos, de los pretratamientos y su orden de aplicación, radica en el efecto que tienen estos sobre el ruido espectral y la dispersión de los datos. Realizando una comparativa entre los espectros de las muestras con y sin aplicación de pretratamientos, se apreció que, mediante la aplicación de los pretratamientos se consiguió de forma general disminuir el ruido espectral y la dispersión. A continuación, se representan los espectros de los datos en crudo que se emplearon para el desarrollo del modelo 1 y los espectros de los datos tras la aplicación de pretratamientos para el desarrollo de los modelos número 8, 9, 12 y 13.

Se han seleccionado los datos de estos 4 modelos para su representación debido a que, tal y como se ha resaltado en la Tabla 5, corresponden con los 4 modelos que mejores resultados de clasificación obtuvieron en la validación cruzada. En la Figura 14 se observan los espectros de las muestras que fueron empleadas para el desarrollo del modelo de calibración sin pretratamientos. En la Figura 15 se muestran los espectros de las muestras tras la aplicación de pretratamientos. La representación de los espectros de las muestras tras la aplicación del resto de pretratamientos, y su combinación, quedan recogidas en el apartado “1. Representación del efecto de los pretratamientos sobre los datos espectrales” del Anexo.

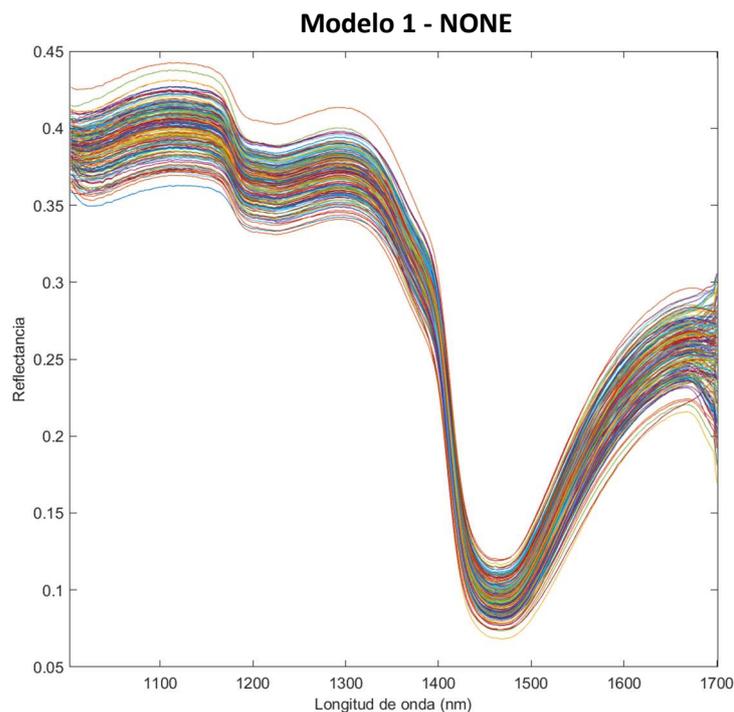


Figura 14. Espectros de las muestras del grupo de calibración en crudo, sin aplicación de pretratamientos (NONE)

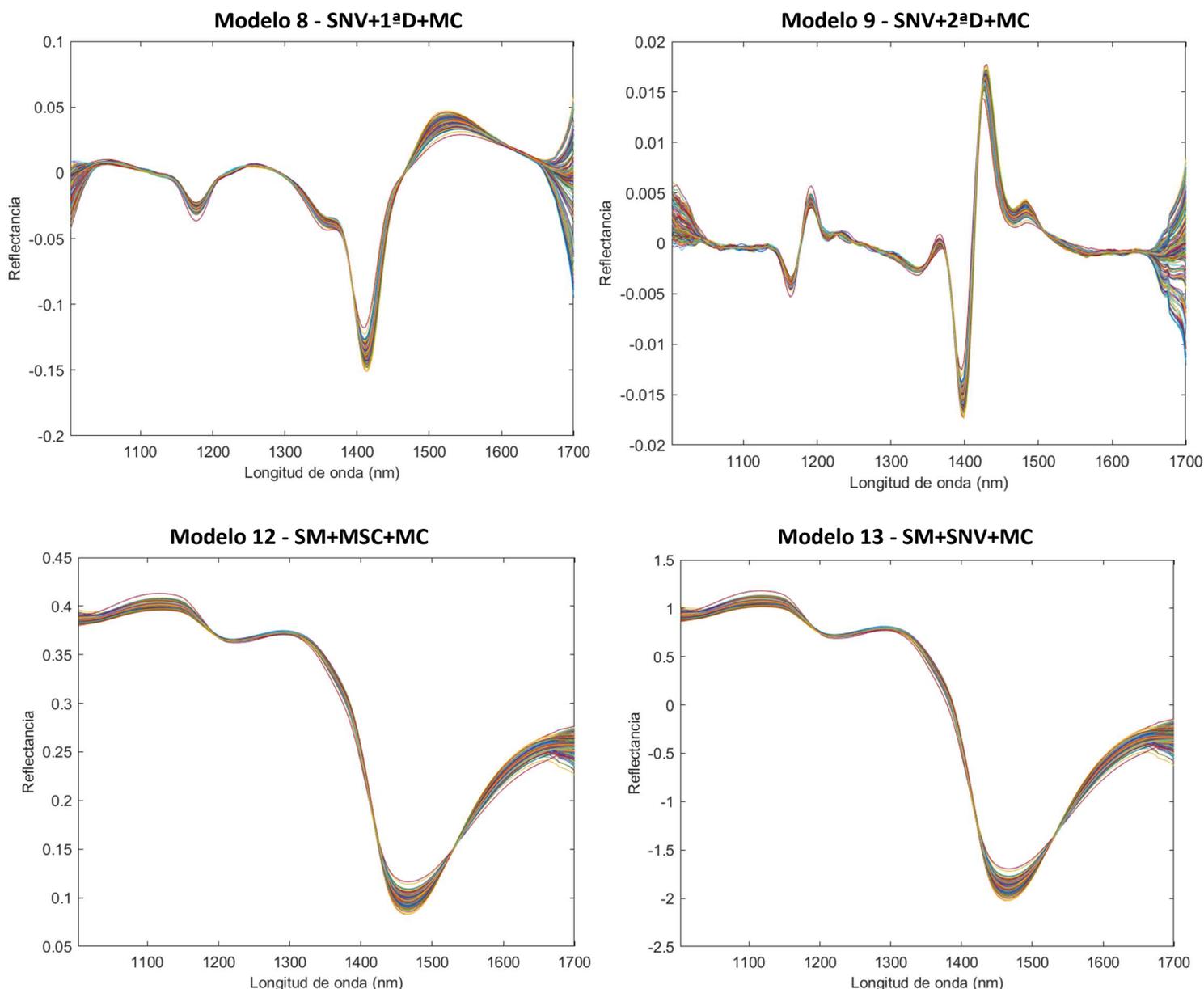


Figura 15. Efecto de los pretratamientos en los espectros de las muestras del grupo de calibración: (MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada)

Comparando los gráficos de la Figura 14 con el gráfico de la Figura 15 se puede detectar el efecto de la aplicación de los pretratamientos sobre los datos espectrales. Mientras que, con la combinación de pretratamientos aplicados en el modelo 8 (SNV+1ªD+MC) y en el modelo 9 (SNV+2ªD+MC) se consiguió una corrección de la dispersión de datos, con los pretratamientos aplicados en el modelo 12 (SM+MSC+MC) y 13 (SM+SNV+MC) se consiguió, además de reducir la dispersión, suavizar los datos y eliminar el ruido espectral.

La selección de los mejores modelos de calibración desarrollados (más precisos) se realizó en función del porcentaje de muestras que cada modelo había sido capaz de clasificar correctamente en sus clases, así como por la sensibilidad, especificidad y el error de clase. Los valores que se obtuvieron en cada modelo de estos 3 últimos parámetros se recogen en la Tabla 6. Los valores de sensibilidad, especificidad, error y porcentaje de muestras correctamente clasificadas, para cada modelo, se

obtuvieron aplicando las Ecuaciones 2, 3, 4 y 5, respectivamente. Estas ecuaciones se describen en el apartado “3.5. Análisis quimiométrico: análisis multivariante”.

Los modelos PLS-DA número 8 (SNV+1ªD+MC), 9 (SNV+2ªD+MC), 12 (SM+MSC+MC) y 13 (SM+SNV+MC) fueron clasificados como los 4 mejores modelos de calibración obtenidos en el presente estudio. En los 4 modelos se aplicó un pretratamiento de normalización ponderada, concretamente, en el modelo número 8, 9 y 13 se aplicó el *Standard Normal Variate* y, en el modelo número 12, el pretratamiento *Multiplicative Scatter Correction*. Cabe destacar que, en todos los modelos se aplicó el pretratamiento *Mean Centering* con el fin de reducir el ruido espectral producido por el sistema.

Tabla 6. Resultados de sensibilidad, especificidad y error de los modelos de calibración desarrollados calculados para la clase “control”. Resaltado, para cada parámetro, los mejores resultados obtenidos (Cal: Calibración; CV: Validación Cruzada; NONE: crudo; MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada)

Modelo	Pretratamiento	Sensibilidad (Cal)	Especificidad (Cal)	Error (Cal)	Sensibilidad (CV)	Especificidad (CV)	Error (CV)
1	NONE	0,590	0,716	0,348	0,578	0,737	0,343
2	MC	0,590	0,723	0,344	0,568	0,681	0,375
3	SNV+MC	0,474	0,819	0,354	0,442	0,766	0,396
4	MSC+MC	0,659	0,755	0,293	0,591	0,713	0,348
5	SM+MC	0,484	0,853	0,332	0,432	0,800	0,384
6	SNV+SM+MC	0,543	0,796	0,331	0,511	0,763	0,363
7	MSC+SM+MC	0,606	0,713	0,340	0,564	0,691	0,372
8	SNV+1ªD+MC	0,547	0,83	0,311	0,537	0,787	0,338
9	SNV+2ªD+MC	0,628	0,763	0,304	0,638	0,71	0,326
10	MSC+1ª D+MC	0,674	0,688	0,319	0,609	0,677	0,357
11	MSC+2ªD+MC	0,649	0,766	0,293	0,585	0,734	0,340
12	SM+MSC+MC	0,632	0,723	0,322	0,621	0,713	0,333
13	SM+SNV+MC	0,685	0,688	0,314	0,663	0,699	0,319

Tras evaluar los modelos en función del % de muestras correctamente clasificadas, la sensibilidad, la especificidad y el error de clase, el modelo de calibración que mejores resultados arrojó fue el número 13. Este modelo se caracteriza por haberse desarrollado tras haber aplicado, en el orden citado, los siguientes pretratamientos a los datos espectrales: (1) SM, (2) SNV y, (3) MC. Cabe mencionar que, fue seleccionado como el modelo de calibración más preciso de entre los 13 desarrollados debido a que se correspondía con el modelo que mejor % CC obtuvo, así como con el modelo que mejores resultados obtuvo de sensibilidad (Cal y CV) y de error de clase en la validación cruzada.

El modelo número 13 fue capaz de clasificar correctamente el 68,65% de las muestras. Cabe destacar que, al igual que en el resto de los modelos PLS-DA desarrollados, las muestras pertenecientes a la clase “déficit de nitrógeno” fueron clasificadas mejor que aquellas pertenecientes a la clase “control”. Concretamente, el modelo número 13 fue capaz de clasificar correctamente un 69,89% de las muestras de la clase “déficit de nitrógeno” y, un 67,39% de las muestras de la clase “control”. Además, la sensibilidad en la calibración y en la validación cruzada alcanzó un valor de 0,685 y 0,663, respectivamente. En cuanto al error de clase en la CV, este fue de 0,319. Sin embargo, a pesar de clasificar el modelo número 13 como el más preciso, no puede ser considerado como un modelo capaz de discriminar la clase “control” de la clase “déficit de nitrógeno”.

Tal y como se menciona en el estudio “clasificación en línea de patatas irregulares mediante un sistema automatizado de visión artificial por ordenador”, un modelo de calibración puede considerarse razonable si el porcentaje arrojado de muestras correctamente clasificadas es  $\geq 75\%$  (Elmasry et al., 2012). En base a ello, ningún modelo de calibración desarrollado en el presente estudio se pudo considerar como efectivo para la clasificación de patatas según los efectos, en hojas, de la fertilización nitrogenada.

Este hecho se pudo haber dado debido a que, pese a realizar un aporte de nitrógeno deficitario en las parcelas con tratamiento “déficit de nitrógeno”, las plantas de patatas pudieron no haber sufrido este estrés abiótico. Esto pudo haber ocurrido debido a diferentes factores:

- La recolección de las muestras (hojas de patata) se realizó en un estado de desarrollo del cultivo temprano (30 días post siembra) previo al pico máximo de la absorción de N del suelo (Alonso, 2002). De esta forma, las plantas, hasta ese momento, pudieron no haber sufrido aún los efectos del déficit de N.
- Las variedades cultivadas bajo estos tratamientos podrían ser poco sensibles a la variación de la cantidad de nitrógeno aportada en la fertilización.
- La cantidad de nitrógeno disponible para las plantas presente en el suelo podría haber sido suficiente para compensar el déficit de N de las parcelas con tratamientos “déficit de nitrógeno”.

En relación con los resultados obtenidos en el presente estudio y, siguiendo las hipótesis generadas, se decidió comprobar la eficacia del empleo de las mismas imágenes hiperespectrales para la clasificación de las hojas de patatas según los efectos de la variedad. Se emplearon los mismos datos espectrales que se utilizaron para el desarrollo de los modelos de calibración destinados a la clasificación de las hojas de patatas en función de los efectos de la fertilización nitrogenada. Con estos, se generó un modelo de calibración PLS-DA diferenciando las clases por variedades en vez de por tratamientos.

Partiendo de que en el presente estudio se seleccionó el modelo número 13 como el modelo de calibración más preciso, se aplicaron, en el mismo orden, los mismos pretratamientos de datos para el desarrollo del modelo de calibración para la clasificación de hojas de patatas en función de los efectos de la variedad. De esta forma, a los datos espectrales recogidos en la matriz de 192 filas por 223 columnas, se les aplicó: (1) *Smoothing*, (2) *Standard Normal Variate* y, (3) *Mean Centering*. Una vez se obtuvieron los datos pretratados, el modelo de calibración se desarrolló de igual manera que los 13 anteriormente generados (véase “3.5. Análisis quimiométrico: análisis multivariante”). Las características del modelo se recogen en la Tabla 7.

Tabla 7. Características del modelo de calibración generado para la clasificación de patatas en función del efecto de la variedad (CC: Correctamente Clasificadas; CV: Validación Cruzada; LV: Variables Latentes; MC: Mean Centering; SNV: Standard Normal Variate; SM: Smoothing)

Nº Muestras	Outliers	Pretratamiento	Var. (%)	LV	% CC (CV)						Grupal
					Alibaba	Baraka	Laudine	Mayka	Romani	Ventana	
173	19	SM+SNV+MC	99,44	4	75,00	72,41	73,33	92,59	66,67	71,43	75,14

Tal y como se predijo, el modelo de calibración desarrollado diferenciando las clases por variedad, en vez de por tratamientos (control/déficit de nitrógeno), fue capaz de clasificar correctamente un 75,14% de las muestras objeto de estudio en la validación cruzada. Esto supone aproximadamente un 6,50% más de muestras correctamente clasificadas en comparación con el resultado de clasificación del modelo número 13 (modelo más preciso del presente estudio). En la Figura 16 se muestra un ejemplo de las predicciones que proporcionó este modelo. Concretamente se ha graficado la predicción para la variedad Mayka correspondiente con la mejor predicción generada (92,59 %CC). Los gráficos de las predicciones restantes quedan recogidos en el apartado “2. Representación de las predicciones obtenidas en la validación cruzada (CV) del modelo PLS-DA para la clasificación de hojas en función del efecto de la variedad” del Anexo.

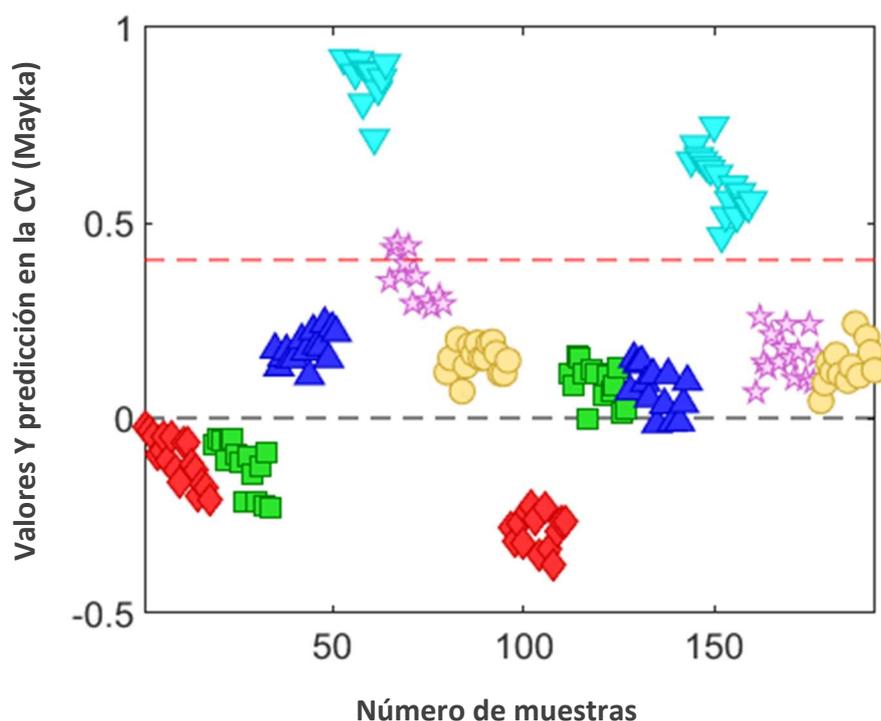


Figura 16. Predicción en la validación cruzada (CV) de la clasificación de las muestras correspondientes con la variedad Mayka)

En definitiva, se ve más variabilidad espectral entre variedades que entre tratamientos. En este caso, según investigadores, se podría considerar que este modelo de calibración es efectivo para la clasificación de patatas según los efectos, en hojas, de la variedad (Elmasry et al., 2012).

#### 4.2. Contraste de resultados con el método de referencia

Con el objetivo de comprobar los resultados obtenidos en el presente estudio, se decidió contrastarlos con los resultados de los rendimientos de cosecha obtenidos por NEIKER de los mismos cultivos de patatas.

Para ello, se realizó un análisis ANOVA con el fin de ver si realmente existían diferencias significativas de rendimiento en función del tipo de tratamiento sabiendo que, el efecto de un estrés abiótico a lo largo del desarrollo del cultivo de patata repercute en el rendimiento de la cosecha. Los estadísticos descriptivos se recogen en la Tabla 8 y los resultados del análisis ANOVA se muestran en la Tabla 9.

Tabla 8. Resultados estadísticos descriptivos del análisis ANOVA de un factor (tratamiento control/déficit de nitrógeno) para la variable rendimiento (kg/ 10 plantas)

Grupos	Nº muestras	Media rendimiento	Desviación estándar	Error estándar
Control	6	15,46	3,17	1,30
Déficit de nitrógeno	6	15,26	2,95	1,20

Tabla 9. Resultados del análisis ANOVA de un factor (tratamiento control/déficit de nitrógeno) para la variable rendimiento (kg/ 10 plantas)

	Suma de cuadrados	Grados de libertad	Media cuadrática	Valor F	Valor P
Entre grupos (control/déficit de N)	0,159	1	0,159	0,017	0,899

En la Tabla 8 se muestra cómo el rendimiento medio de la cosecha de las plantas cultivadas en las parcelas sometidas a la clase de tratamiento “control” fue ligeramente superior a los de la clase “déficit de nitrógeno”. Sin embargo, tal y como se observa en la Tabla 9, la significancia asintótica o “Valor p” resultó ser mayor a 0,05 (nivel de significancia con el que convencionalmente se trabaja) en el análisis ANOVA. De esta manera, se aceptó la hipótesis nula (hipótesis de igualdad) y concluimos que no existieron diferencias estadísticamente significativas en el rendimiento medio de la cosecha entre clases de tratamientos.

En base al resultado del análisis ANOVA, determinamos que la recolección de hojas en un estado temprano del desarrollo del cultivo no fue el factor por el que las plantas no sufrieron estrés dado que las plantas obtuvieron rendimientos iguales bajo ambas condiciones de tratamientos (control/déficit de nitrógeno). De esta manera, cabe la posibilidad de que (1) las variedades cultivadas bajo estos tratamientos fuesen poco sensibles a la variación de la cantidad de nitrógeno aportada o, (2) la cantidad de N disponible para las plantas presente en el suelo fue suficiente para compensar el déficit de N.

Quedando fuera del objetivo del proyecto, pero debido a su gran relevancia, se recomienda continuar profundizando en la presente investigación tanto para conocer el factor determinante de las respuestas de las plantas ante los 2 tratamientos de fertilización, como para conocer si las variedades cultivadas son poco sensibles y, por ende, resistentes a un déficit de N. Este último aspecto es un punto muy relevante para tenerlo en cuenta, tanto desde el punto de vista económico como medioambiental, ya que el empleo de estas variedades permitiría utilizar dosis de nitrógeno inferiores suponiendo un beneficio económico y medioambiental.

### 4.3. Validación externa de los modelos de calibración

Antes de recomendar el uso práctico de un modelo predictivo es importante que este sea nuevamente evaluado en una muestra de datos independientes a la muestra de datos utilizada para su desarrollo (Knottnerus, 1995). Por ello es por lo que, teóricamente, la validación externa se realiza únicamente sobre el mejor modelo desarrollado para un estudio dado.

Teniendo en cuenta que, para el presente estudio, ningún modelo de calibración desarrollado se ha podido considerar como efectivo, no hubiese hecho falta realizar una validación externa ya que probablemente nunca llegarían a ser puestos en práctica. Sin embargo, se continuó con el proceso de validación externa con el fin de aplicar el protocolo completo de investigación del presente estudio y corroborar la baja precisión de los modelos PLS-DA generados.

Así, una vez desarrollados los modelos de calibración se volvió a evaluar la precisión de cada uno de ellos a través de una validación externa con el fin de comprobar si funcionaban igual con otros datos diferentes a los que se habían empleado en el desarrollo de los mismos (Altman y Royston, 2000). Para ello, cada modelo de calibración fue validado externamente con el 20% de las muestras (48) destinadas inicialmente para este fin y, por ende, desconocidas para los modelos desarrollados tal y como se explica en el apartado "3.4. Procesado de imágenes". Los datos empleados para la validación externa se agruparon en una matriz la cual recogía toda la información espectral de las 48 muestras, es decir, 24 muestras de cada clase (control/déficit de nitrógeno).

Debido a que los datos espectrales presentaban mucho ruido, con el fin de disminuirlo, se aplicaron 2 pretratamientos previo a realizar la validación externa: (1) suavizado con SM (*Smoothing Savitzky-Golay*) y, (2) normalización ponderada con SNV (*Standard Normal Variate*). En primer lugar, se aplicó únicamente un *Smoothing* pero tras ver los resultados se decidió aplicar también un SNV. En la Figura 17 se muestra cómo tras la aplicación del pretratamiento SM se consiguió minimizar el ruido espectral y, que, tras la aplicación del SM y del SNV, en combinación, se consiguió además de reducir el ruido espectral, reducir la dispersión de los datos.

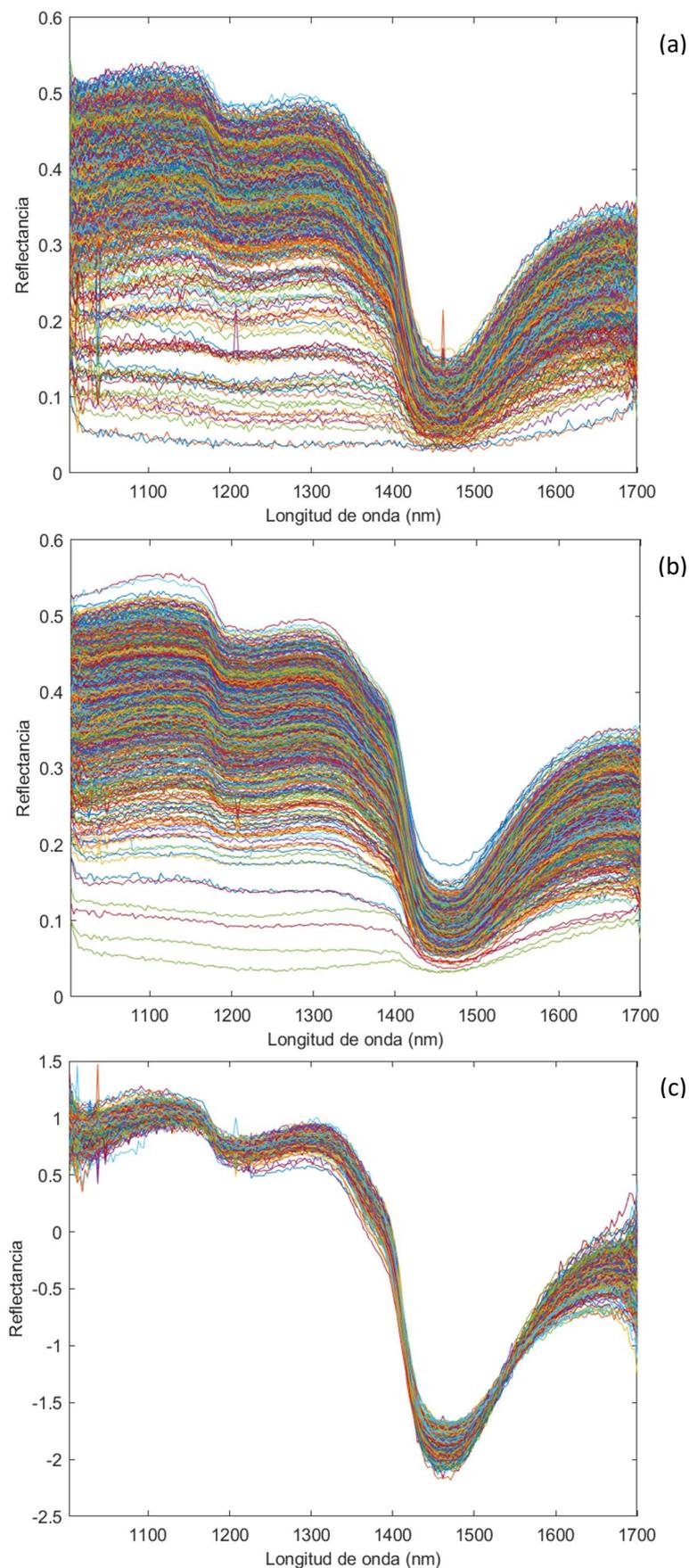


Figura 17. Efecto de los pretratamientos en los espectros de las muestras del grupo de validación: a) NONE, b) SM y c) SM+SNV (NONE: Sin pretratamiento; SM: Smoothing; SNV: Standard Normal Variate)

En la Tabla 10 se muestran, para cada modelo de calibración desarrollado, los resultados que se obtuvieron en la validación externa respecto al % de muestras correctamente clasificadas. Así mismo, en la Tabla 10 se recoge: (1) número de modelo de calibración y (2) pretratamientos aplicados a los datos espectrales en cada modelo de calibración, (3) número de muestras empleadas para la validación externa del modelo, (4) número de *outliers*, (5) porcentaje de muestras correctamente clasificadas (CC) en la validación externa por cada clase y, grupalmente.

Cabe mencionar que, el número de muestras eliminadas y, por ende, no consideradas para la validación externa del modelo de calibración en cuestión, supuso una pérdida de datos inferior al 10% del total.

Tabla 10. Porcentaje de muestras correctamente clasificadas (%CC) en la validación externa de cada modelo de calibración desarrollado en el presente estudio. Resaltados los modelos más precisos obtenidos tras la validación externa. (NONE: crudo; MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada)

Modelo	Pretratamiento	Nº muestras	Outliers	%CC (Validación externa)		
				Control	Déficit de nitrógeno	Grupal
1	NONE	111227	4606	0,00%	100,00%	49,73%
2	MC	114659	1174	0,32%	100,00%	49,65%
3	SNV+MC	107100	8733	89,34%	11,30%	50,59%
4	MSC+MC	112173	3660	50,20%	52,29%	51,68%
5	SM+MC	112173	3660	100,00%	0,00%	50,25%
6	SNV+SM+MC	106331	9502	58,46%	51,67%	54,98%
7	MSC+SM+MC	111882	3951	52,66%	50,33%	51,53%
8	SNV+1ªD+MC	112734	3099	48,21%	55,94%	52,04%
9	SNV+2ªD+MC	114302	1531	50,57%	51,30%	50,93%
10	MSC+1ª D+MC	111568	4265	51,28%	52,69%	51,97%
11	MSC+2ªD+MC	108248	7585	52,80%	52,52%	52,66%
12	SM+MSC+MC	104603	11230	52,03%	52,81%	52,41%
13	SM+SNV+MC	110628	5205	54,00%	52,09%	53,06%

Teóricamente, el mayor porcentaje de muestras correctamente clasificadas en la validación externa se debería haber obtenido al validar el modelo de calibración desarrollado más preciso, en este caso, el modelo de calibración número 13 (véase “4.1. Modelos de calibración para clasificar patatas según los efectos de la fertilización”). Sin embargo, en el presente estudio, el valor del %CC en la validación externa del modelo de calibración 13 se situó en segundo lugar siendo la validación externa del modelo número 6 la que mejores resultados de %CC arrojó (Tabla 10). El modelo número 6 (SNV+SM+MC) fue capaz de clasificar correctamente un 54,98% de las muestras mientras que el modelo número 13 (SM+SNV+MC) clasificó correctamente un 53,06% de las muestras objeto de estudio en la validación externa.

A pesar de ello, la diferencia entre ambos se consideró insignificante siendo esta de un 1,92% y, teniendo en cuenta los resultados, se comprobó la baja precisión y eficacia de los modelos de calibración desarrollados. En definitiva, los modelos en cuestión realizaron una clasificación prácticamente al azar, de las hojas de patatas, en función del efecto de la fertilización nitrogenada.

En la Figura 18 se muestra la representación gráfica, para los modelos 6 y 13, de la matriz de confusión para cada grupo incluido en el estudio. La línea roja horizontal de puntos indica el umbral por encima del cual una muestra se asigna a una clase determinada, en este caso, la clase “control”. Por lo tanto, en la Figura 18, todas las muestras situadas por encima del umbral se clasificaron como pertenecientes al grupo “control” y, todas las muestras por debajo del umbral se clasificaron como pertenecientes al grupo “déficit de nitrógeno”. Se puede observar cómo, efectivamente, los modelos de calibración realizaron una clasificación casi al azar de las muestras destinadas a la validación externa.

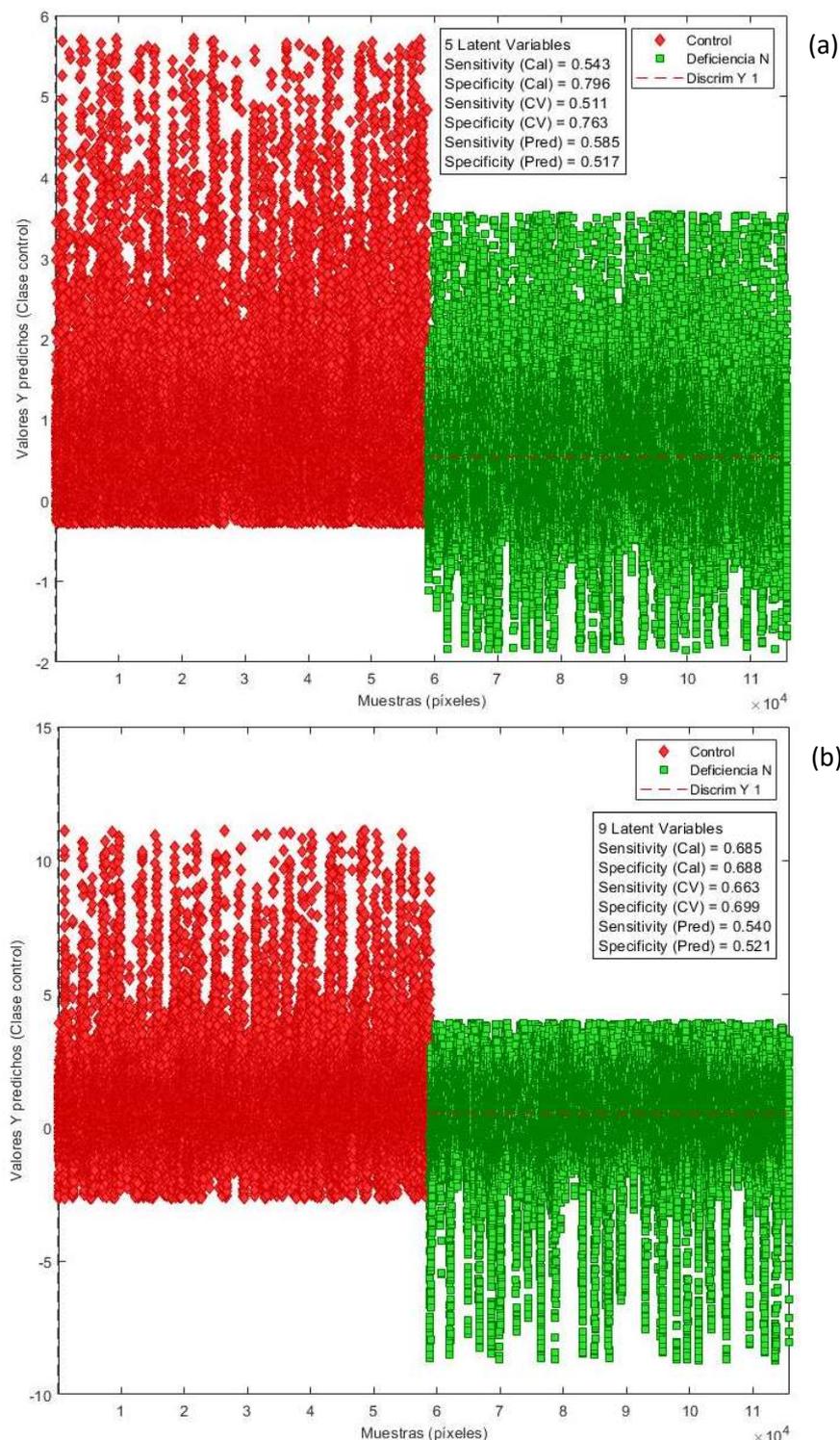


Figura 18. Representación gráfica de los resultados de predicción que ejecuta el modelo (a) número 6 y, (b) número 13, sobre la matriz de validación externa

En la Tabla 11 se resumen los valores de sensibilidad, especificidad y error que se obtuvieron en la validación externa de los modelos PLS-DA número 6 y 13. Dichos resultados fueron los mejores obtenidos en la validación externa de los modelos de calibración desarrollados en el presente estudio. Como se muestra en la Tabla 11, para ambos modelos validados externamente, el valor de sensibilidad de la clase “control” fue mayor que el de la clase “déficit de nitrógeno”. Sin embargo, siendo los valores de sensibilidad y de especificidad inferiores a 0,60, los errores de clase igual o superior a 0,45 y, teniendo en cuenta los %CC obtenidos en la validación externa (Tabla 10), no se pudieron catalogar los modelos como precisos. Así, a través de la validación externa, se comprueba una vez más la baja eficiencia de los modelos PLS-DA desarrollados para clasificar las hojas de patatas objeto de estudio en función de los tratamientos de fertilización empleados.

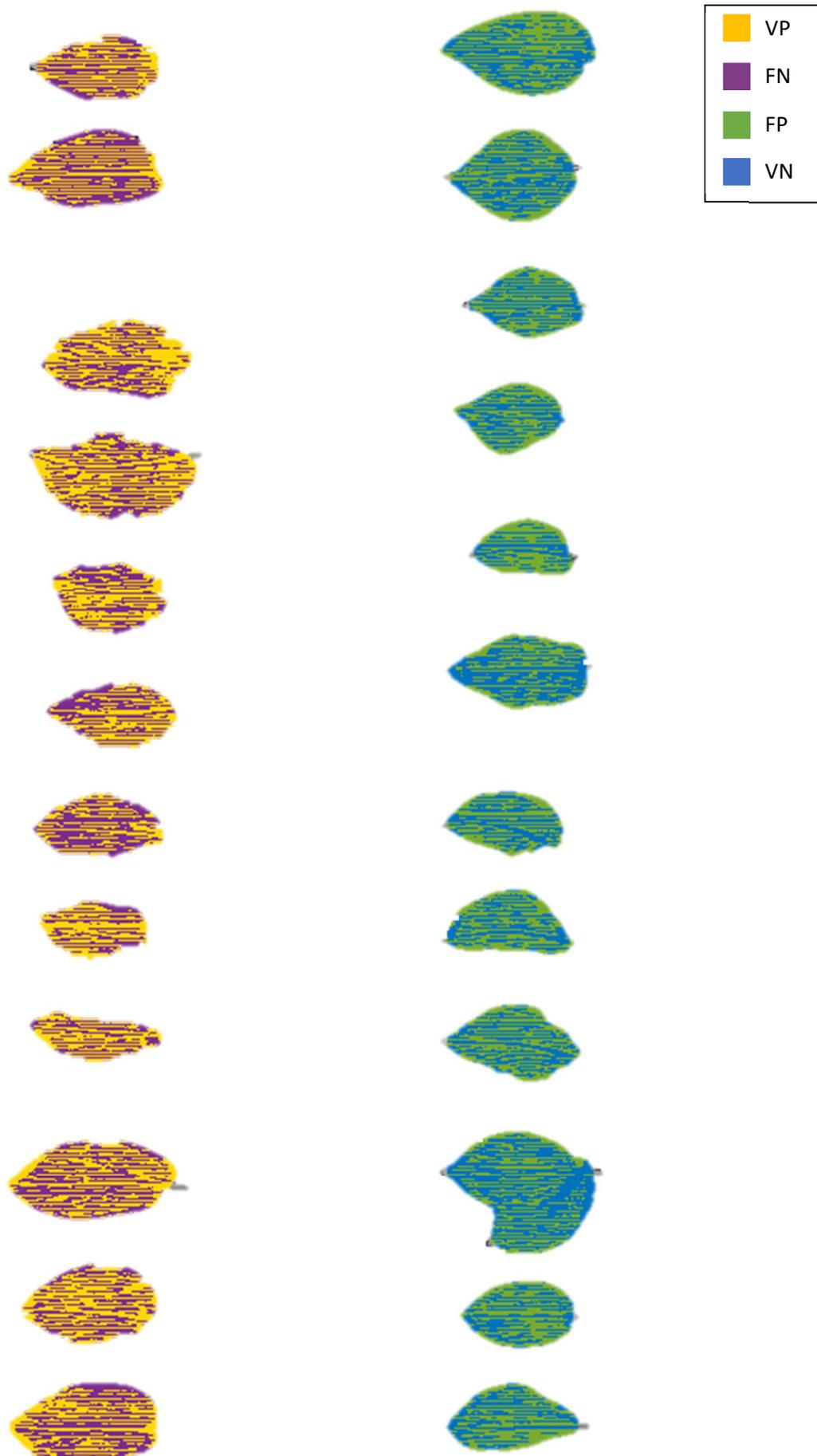
Tabla 11. Resultados de la validación externa del modelo de calibración número 6 y 13 (Pred: predicción; MC: Mean Centering; SNV: Standard Normal Variate; SM: Smoothing)

Modelo	Pretratamientos	Clase	Sensibilidad (Pred)	Especificidad (Pred)	Error (Pred)
6	SNV+SM+MC	Control	0,585	0,517	0,449
		Déficit de nitrógeno	0,517	0,585	0,449
13	SM+SNV+MC	Control	0,540	0,521	0,470
		Déficit de nitrógeno	0,521	0,540	0,470

Tras haber realizado la validación externa de los modelos PLS-DA desarrollados, y, gracias a haber empleado para este proceso los datos del espectro completo de cada una de las 48 muestras, se generó para cada una de ellas su imagen química (Figura 19). La imagen química, además de permitir visualizar aquellos píxeles correctamente clasificados en cada muestra, permite detectar qué regiones presentan más complicaciones para ser clasificadas correctamente por los modelos de calibración. Si las regiones donde se concentran píxeles incorrectamente clasificados siguen un patrón de localización (por ejemplo: píxeles del contorno de la hoja) podrían desarrollarse nuevas investigaciones enfocadas hacia el estudio de estas regiones en comparación con el resto de la muestra con el fin de detallar las conclusiones del estudio.

De esta forma, pese a los resultados obtenidos de clasificación, se realizó la imagen química con el fin de visualizar los píxeles correctamente clasificados y detectar, en el caso de que lo hubiera, regiones específicas de mayor concentración de píxeles clasificados incorrectamente. Así, se obtuvo un mapa representativo de cada una de las 48 muestras (hojas de patata) empleadas en la validación externa (Figura 19).

Tal y como se observa en la Figura 19, los píxeles incorrectamente clasificados en su clase no siguen un patrón de localización, simplemente, tal y como se ha comprobado, los modelos PLS-DA generados clasifican prácticamente al azar los datos objeto de estudio.



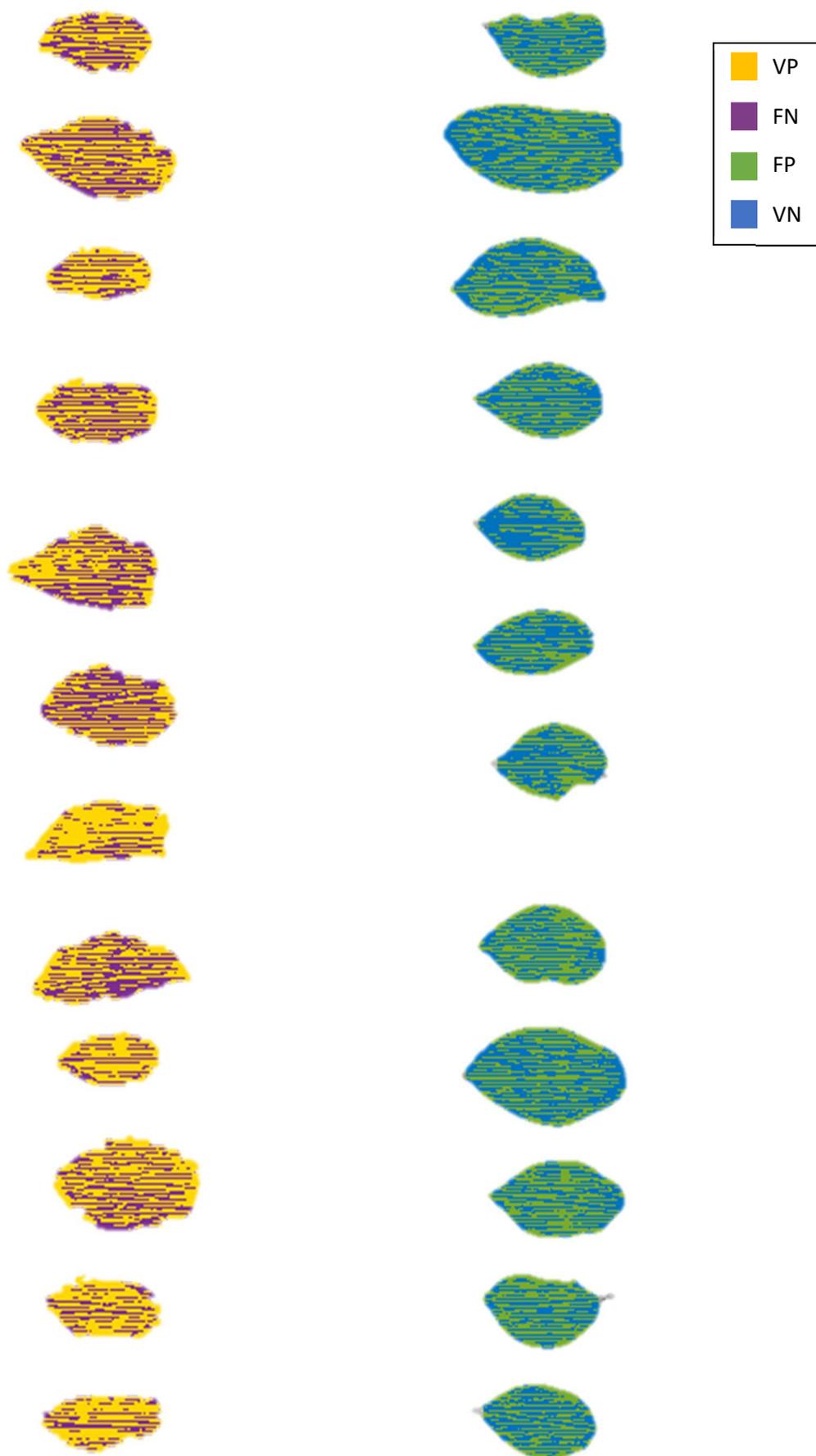


Figura 19. Imagen química de las 48 muestras empleadas para la validación externa de los modelos PLS-DA desarrollados

## 5. CONCLUSIONES

- Los mejores resultados de calibración se obtienen con 3 modelos en los que se ha aplicado el pretratamiento *Standard Normal Variate* en los datos espectrales, consiguiendo los 3 un porcentaje de muestras correctamente clasificadas superior al 67%.
- El orden de aplicación de los pretratamientos influye, entorno a un 5%, en la capacidad de los modelos PLS-DA desarrollados de clasificar correctamente las muestras objeto de estudio.
- El modelo PLS-DA desarrollado más preciso clasifica bien el 67,39% de las muestras de la clase “control” y el 69,89% de la clase “déficit de nitrógeno” en la validación cruzada.
- El modelo PLS-DA generado más preciso se caracteriza por ser desarrollado a partir de datos espectrales pretratados con los siguientes pretratamientos citados en orden de aplicación: (1) *Smoothing*, (2) *Standard Normal Variate*, (3) *Mean Center*.
- El porcentaje de muestras correctamente clasificadas por el modelo de calibración más preciso, en la validación externa, es de 53,06%. Por ello, se determina que los modelos en cuestión clasifican al azar las hojas de patata (objeto de estudio) en función del efecto de la fertilización nitrogenada.
- Los píxeles incorrectamente clasificados en su clase no siguen un patrón de localización concreto según se muestra en las imágenes químicas de las muestras destinadas para la validación externa de los modelos PLS-DA desarrollados.
- Existe más variabilidad espectral entre hojas de plantas de patata de diferentes variedades que, entre hojas de plantas sometidas a diferentes tratamientos de fertilización nitrogenada.
- Las variedades sometidas a un déficit de nitrógeno responden de igual manera que las cultivadas en condiciones óptimas de este nutriente. Se recomienda investigar acerca de este hecho ya que las variedades empleadas en el presente estudio podrían ser tolerantes al déficit de nitrógeno y, por ende, interesantes para incluirlas en programas de mejora genética futuros con el fin de obtener beneficios, a largo plazo, tanto medioambientales como económicos.
- La baja precisión de clasificación de los modelos PLS-DA desarrollados se debe a que, ciertamente, no han existido diferencias, ni físicas ni químicas, entre las hojas de las plantas de patata cultivadas bajo un déficit de nitrógeno y las cultivadas en condiciones óptimas de este nutriente.
- La técnica de imágenes hiperespectrales (HSI) en el rango infrarrojo cercano (NIR) no se ha mostrado efectiva como herramienta para identificar la respuesta, en las hojas de patata estudiadas, a la deficiencia de abonado nitrogenado ya que ningún modelo de calibración desarrollado se ha podido considerar preciso para este fin.

## 6. BIBLIOGRAFÍA

- Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air CO<sub>2</sub> enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO<sub>2</sub>. *New Phytologist*, 165(2), 351–372. <https://doi.org/https://doi.org/10.1111/j.1469-8137.2004.01224.x>
- Alonso, F. (2002). *El cultivo de la patata*. Ediciones Mundi Prensa. Retrieved from <https://bvirtual.uce.edu.ec:2352/lib/ucesp/reader.action?docID=3175751&query=CULTIVO+DE+PAPAS+>
- Altman, D., & Royston, P. (2000). Altman DG, Royston P What do we mean by validating a prognostic model? *Stat Med* 19: 453-473. *Statistics in Medicine*, 19, 453–473. [https://doi.org/10.1002/\(SICI\)1097-0258\(20000229\)19:43.3.CO;2-X](https://doi.org/10.1002/(SICI)1097-0258(20000229)19:43.3.CO;2-X)
- Amigo, J. M., Babamoradi, H., & Elcoroaristizabal, S. (2015). Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta*, 896, 34–51. <https://doi.org/https://doi.org/10.1016/j.aca.2015.09.030>
- Atherton, D., Watson, D. G., Zhang, M., Qin, Z., & Liu, X. (2015). Hyperspectral Spectroscopy for Detection of Early Blight (*Alternaria solani*) Disease in Potato (*Solanum tuberosum*) Plants at Two Different Growth Stages. In *2015 ASABE Annual International Meeting* (p. 1). American Society of Agricultural and Biological Engineers.
- Casa, A. de la, Ovando, G., Bressanini, L., & Martínez, J. (2013). Estimación de la cobertura del suelo y evaluación de la fertilización nitrogenada en papa con registros de cámaras digitales convencionales y modificadas. In *V Congreso Argentino de AgroInformática (CAI)-JAIIO 42 (2013)*.
- Castro, W. M. (2015). *APLICACIÓN DE LA TECNOLOGÍA DE IMÁGENES HIPERESPECTRALES AL CONTROL DE CALIDAD DE PRODUCTOS AGROALIMENTARIOS DE LA REGIÓN DE AMAZONAS (PERÚ)*. Universidad Politécnica de Valencia. Retrieved from <https://m.riunet.upv.es/bitstream/handle/10251/63250/-CASTRO - Aplicación de la tecnología de imágenes hiperespectrales al control de calidad de produc....pdf?sequence=1&isAllowed=y>
- CCAE. (2004). Caracterización del mercado y de la comercialización de la patata en las cooperativas agrarias. Retrieved from [https://www.mapa.gob.es/es/alimentacion/temas/consumo-tendencias/patata\\_tcm30-89316.pdf](https://www.mapa.gob.es/es/alimentacion/temas/consumo-tendencias/patata_tcm30-89316.pdf)
- Ceballos, M. del C. J., Roncal, C. P., García, R. P., Maestresalas, A. L., & de Galarreta, J. I. R. (2019). Evaluación de las pérdidas en patata durante su almacenamiento por punción y tecnología NIRS. In *X Congreso Ibérico de Agroingeniería: Libro de actas* (pp. 1221–1226). Universidad de Zaragoza. Retrieved from <https://zaguan.unizar.es/record/84480/files/3438.pdf>
- Comisión Europea. (2021a). Causas del cambio climático. Retrieved October 14, 2021, from [https://ec.europa.eu/clima/climate-change/causes-climate-change\\_es](https://ec.europa.eu/clima/climate-change/causes-climate-change_es)
- Comisión Europea. (2021b). Consecuencias del cambio climático. Retrieved October 14, 2021, from [https://ec.europa.eu/clima/climate-change/climate-change-consequences\\_es](https://ec.europa.eu/clima/climate-change/climate-change-consequences_es)
- Dacal-Nieto, A., Formella, A., Carrión, P., Vazquez-Fernandez, E., & Fernández-Delgado, M. (2011a). Common scab detection on potatoes using an infrared hyperspectral imaging system. In *International conference on image analysis and processing* (pp. 303–312). Springer.
- Dacal-Nieto, A., Formella, A., Carrión, P., Vazquez-Fernandez, E., & Fernández-Delgado, M. (2011b). Non-destructive detection of hollow heart in potatoes using hyperspectral imaging. In *International Conference on Computer Analysis of Images and Patterns* (pp. 180–187). Springer.

- de Almeida, M. R., Correa, D. N., Rocha, W. F. C., Scafi, F. J. O., & Poppi, R. J. (2013). Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation. *Microchemical Journal*, *109*, 170–177. <https://doi.org/https://doi.org/10.1016/j.microc.2012.03.006>
- Deryng, D., Elliott, J., Folberth, C., Müller, C., Pugh, T. A. M., Boote, K. J., ... Rosenzweig, C. (2016). Regional disparities in the beneficial effects of rising CO<sub>2</sub> concentrations on crop water productivity. *Nature Climate Change*, *6*(8), 786–790. <https://doi.org/10.1038/nclimate2995>
- Dorrepaal, R., Malegori, C., & Gowen, A. (2016). Tutorial: Time Series Hyperspectral Image Analysis. *J. Near Infrared Spectrosc.*, *24*(2), 89–107. Retrieved from <http://www.osapublishing.org/jnirs/abstract.cfm?URI=jnirs-24-2-89>
- Easlson, H. M., & Bloom, A. J. (2013). The effects of rising atmospheric carbon dioxide on shoot-root nitrogen and water signaling. *Frontiers in Plant Science*, *4*, 304. Retrieved from <https://www.frontiersin.org/articles/10.3389/fpls.2013.00304/full>
- EEA. (2016). Climate change, impacts and vulnerability in Europe 2016 - An indicator-based report. *EEA Report*, (1/2017). Retrieved from <https://www.eea.europa.eu/publications/climate-change-adaptation-and-disaster>
- EEA. (2019). Climate change adaptation in the agriculture sector in Europe. *EEA Report*, (04/2019), 112.
- EEA. (2020a). Climate change threatens future of farming in Europe. Retrieved October 14, 2021, from <https://www.eea.europa.eu/highlights/climate-change-threatens-future-of/#Figure>
- EEA. (2020b). La adaptación al cambio climático es clave para el futuro de la agricultura en Europa. Retrieved October 14, 2021, from <https://www.eea.europa.eu/es/articulos/la-adaptacion-al-cambio-climatico>
- EEA. (2021). Agriculture and climate change. Retrieved from <https://www.eea.europa.eu/signals/signals-2015/articles/agriculture-and-climate-change>
- Elmasry, G., Cubero, S., Molto, E., & Blasco, J. (2012). In-line sorting of irregular potatoes by using automated computer-based machine vision system. *Journal of Food Engineering*, *112*, 60–68. <https://doi.org/10.1016/j.jfoodeng.2012.03.027>
- Elmasry, G., Kamruzzaman, M., Sun, D.-W., & Allen, P. (2012). Principles and Applications of Hyperspectral Imaging in Quality Evaluation of Agro-Food Products: A Review. *Critical Reviews in Food Science and Nutrition*, *52*(11), 999–1023. <https://doi.org/10.1080/10408398.2010.543495>
- ElMasry, G., & Sun, D.-W. (2010). CHAPTER 1 - Principles of Hyperspectral Imaging Technology. In D.-W. B. T.-H. I. for F. Q. A. and C. Sun (Ed.) (pp. 3–43). San Diego: Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-374753-2.10001-2>
- Eurostat. (2020). Agricultural production - crops - Statistics Explained. Retrieved February 10, 2022, from [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agricultural\\_production\\_-\\_crops#Potatoes\\_and\\_sugar\\_bee](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agricultural_production_-_crops#Potatoes_and_sugar_bee)
- FAO. (2008). Las papas, la nutrición y la alimentación - Año Internacional de la Papa 2008. Retrieved February 10, 2022, from <https://www.fao.org/potato-2008/es/lapapa/hojas.html>
- FAO. (2015a). Climate change and food security: risks and responses. Retrieved from <https://www.fao.org/3/i5188e/i5188e.pdf>

- FAO. (2015b). *Estimación de emisiones de gases de efecto invernadero en la agricultura Un manual para abordar los requisitos*.
- FAO. (2020). FAOSTAT. Retrieved February 10, 2022, from [https://www.fao.org/faostat/es/#rankings/countries\\_by\\_commodity](https://www.fao.org/faostat/es/#rankings/countries_by_commodity)
- FH Joanneum. (2005). *Cross-Validation Explained*. Institute for Genomics and Bioinformatics - Graz University of Technology. Retrieved from <http://genome.tugraz.at/proclassify/help/pages/XV.html>
- García, E., & Flego, F. (2008). Agricultura de precisión. *Revista Ciencia Y Tecnología*.
- George, T. S., Taylor, M. A., Dodd, I. C., & White, P. J. (2017). Climate change and consequences for potato production: a review of tolerance to emerging abiotic stress. *Potato Research*, 60(3), 239–268. Retrieved from <https://link.springer.com/article/10.1007/s11540-018-9366-3>
- Gómez-Sanchis, J., Moltó, E., Camps-Valls, G., Gómez-Chova, L., Aleixos, N., & Blasco, J. (2008). Automatic correction of the effects of the light source on spherical objects. An application to the analysis of hyperspectral images of citrus fruits. *Journal of Food Engineering*, 85(2), 191–200. <https://doi.org/https://doi.org/10.1016/j.jfoodeng.2007.06.036>
- Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G., & Frias, J. M. (2007). Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, 18(12), 590–598. <https://doi.org/https://doi.org/10.1016/j.tifs.2007.06.001>
- Grahn, H. F., & Geladi, P. (2007). *Techniques and Applications of Hyperspectral Image Analysis*. Retrieved from <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470010884>
- Ibárcena Escudero, M., & Scheelje Bravo, J. M. (2003). *El cambio climático principales causantes, consecuencias y compromisos de los países involucrados*. Retrieved from <https://www.fao.org/3/xii/0523-b2.htm>
- International Potato Center. (2017). Hechos y cifras sobre la papa. Lima (Perú). CIP. Retrieved from <https://cgspace.cgiar.org/bitstream/handle/10568/87957/CIP-Hechos-y-cifras-sobre-la-papa-Espanol-2017.pdf?sequence=1&isAllowed=y>
- IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Grupo Intergubernamental de Expertos Sobre El Cambio Climático (IPCC)*. Retrieved from [https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC\\_AR6\\_WGI\\_Full\\_Report.pdf](https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf)
- Ivorra Martínez, E. (2015). *Desarrollo de técnicas de visión hiperespectral y tridimensional para el sector agroalimentario*. Universidad Politécnica de Valencia. Retrieved from [https://riunet.upv.es/bitstream/handle/10251/48541/Ivorra - Desarrollo de técnicas de visión hiperespectral y tridimensional para el sector agroali....pdf?sequence=1](https://riunet.upv.es/bitstream/handle/10251/48541/Ivorra_-_Desarrollo_de_técnicas_de_visión_hiperespectral_y_tridimensional_para_el_sector_agroali....pdf?sequence=1)
- J., B. A., Martin, B., Rubio, A. J. S., & B., C. A. (2010). Carbon Dioxide Enrichment Inhibits Nitrate Assimilation in Wheat and Arabidopsis. *Science*, 328(5980), 899–903. <https://doi.org/10.1126/science.1186440>
- Knottnerus, J. A. (1995). Diagnostic prediction rules: principles, requirements and pitfalls. *Primary Care*, 22(2), 341–363.
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation*. John Wiley & Sons.

- López-Maestresalas, A., Keresztes, J. C., Goodarzi, M., Arazuri, S., Jarén, C., & Saeys, W. (2016). Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging. *Food Control*, 70, 229–241.
- Lopez Maestresalas, A. (2016). *Near-infrared spectroscopy and hyperspectral imaging for non-destructive quality inspection of potatoes*. Universidad Pública de Navarra. Retrieved from [https://academica-e.unavarra.es/bitstream/handle/2454/28669/Tesis\\_Ainara\\_Lopez\\_MA.pdf?sequence=1&isAllowed=y](https://academica-e.unavarra.es/bitstream/handle/2454/28669/Tesis_Ainara_Lopez_MA.pdf?sequence=1&isAllowed=y)
- López Maestresalas, A., Pérez Roncal, C., Jarén, C., Ruiz de Galarreta, J. I., Álvarez, A., Barandalla, L., & Arazuri, S. (2021). Imágenes hiperespectrales para el estudio de la respuesta a los estreses abióticos (deficiencia de riego y abonado) de distintos cultivares de patata. *Horticultura*. Retrieved from [https://www.interempresas.net/Horticola/Articulos/373909-Imagenes-hiperespectrales-estudio-respuesta-estreses-abioticos-\(deficiencia-riego.html](https://www.interempresas.net/Horticola/Articulos/373909-Imagenes-hiperespectrales-estudio-respuesta-estreses-abioticos-(deficiencia-riego.html)
- Loyola, N., Oyarce, E., & Acuña, C. (2010). Evaluación del contenido de almidón en papas (*Solanum tuberosum* Cv. desirée), producidas en forma orgánica y convencional, en la provincia de Curicó, región del Maule. *Idesia (Arica)*, 28(2), 41–52. Retrieved from [https://scielo.conicyt.cl/scielo.php?pid=S0718-34292010000200005&script=sci\\_arttext&lng=en](https://scielo.conicyt.cl/scielo.php?pid=S0718-34292010000200005&script=sci_arttext&lng=en)
- Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., & Kaufman, L. (1988). *Chemometrics: a textbook* (Volume 2). Amsterdam: Elsevier. Retrieved from <https://www.sciencedirect.com/bookseries/data-handling-in-science-and-technology/vol/2/suppl/C>
- MITECO. (2021a). Avance de Emisiones de Gases de Efecto Invernadero correspondientes al año 2020: nota informativa. Retrieved from [https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-sei/avance-gei-2020\\_tcm30-528804.pdf](https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-sei/avance-gei-2020_tcm30-528804.pdf)
- MITECO. (2021b). Inventario nacional de emisiones de gases de efecto invernadero: informe resumen. Retrieved from [https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-sei/documentoresumeninventariogei-ed2021\\_tcm30-524841.pdf](https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/sistema-espanol-de-inventario-sei/documentoresumeninventariogei-ed2021_tcm30-524841.pdf)
- Mollazade, K., Omid, M., Tab, F. A., & Mohtasebi, S. S. (2012). Principles and Applications of Light Backscattering Imaging in Quality Evaluation of Agro-food Products: a Review. *Food and Bioprocess Technology*, 5(5), 1465–1485. <https://doi.org/10.1007/s11947-012-0821-x>
- Municio, D. (2011). *Técnicas de oversampling aplicadas al análisis de imágenes hiperespectrales*. Universidad de extremadura. Retrieved from [https://dehesa.unex.es/bitstream/10662/8811/1/TFGUEx\\_2019\\_Municio\\_Duran.pdf](https://dehesa.unex.es/bitstream/10662/8811/1/TFGUEx_2019_Municio_Duran.pdf)
- NASA. (2016). Rising Carbon Dioxide Levels Will Help and Hurt Crops. Retrieved October 14, 2021, from <https://www.nasa.gov/feature/goddard/2016/nasa-study-rising-carbon-dioxide-levels-will-help-and-hurt-crops>
- NCEZID. (2019). Métodos de transmisión | Virus del Zika | CDC. *Centros Para El Control Y La Prevención de Enfermedades, Centro Nacional de Enfermedades Infecciosas Zoonóticas Y Emergentes (NCEZID)*. Retrieved from <https://www.cdc.gov/zika/es/prevention/transmission-methods.html>
- Oliva Lobo, G. A. (2020). *CLASIFICACIÓN DE PAPA DE ACUERDO CON SU APTITUD DE TRASFORMACIÓN INDUSTRIAL FRITO Y COCIDO USANDO IMÁGENES HIPERESPECTRALES*. Universidad Pública de Navarra. Retrieved from [https://academica-e.unavarra.es/xmlui/bitstream/handle/2454/37955/TFM\\_Oliva\\_Gil.pdf?sequence=1&isAllowed=y](https://academica-e.unavarra.es/xmlui/bitstream/handle/2454/37955/TFM_Oliva_Gil.pdf?sequence=1&isAllowed=y)

- ONU. (2020a). Climate Change and COVID-19: UN urges nations to “recover better.” Retrieved October 14, 2021, from <https://www.un.org/en/un-coronavirus-communications-team/un-urges-countries-“build-back-better”>
- ONU. (2020b). Fall in COVID-linked carbon emissions won’t halt climate change. Retrieved October 14, 2021, from <https://news.un.org/en/story/2020/04/1062332>
- ONU. (2020c). First Person: COVID-19 is not a silver lining for the climate. Retrieved October 14, 2021, from <https://news.un.org/en/story/2020/04/1061082>
- ONU. (2021a). Objetivo 13: Acción por el clima. Retrieved October 14, 2021, from <https://www.un.org/sustainabledevelopment/es/climate-change-2/>
- ONU. (2021b). Objetivos de desarrollo sostenible. Retrieved October 14, 2021, from <https://www.un.org/sustainabledevelopment/es/>
- Rabadan-Diehl, C. (2017). Cambio climático: Nuevas enfermedades y el resurgir de otras casi erradicadas. Retrieved October 14, 2021, from <https://www.efesalud.com/rabadan-diehl-cambio-climatico-enfermedades/>
- Solana, J. (1994). Instrumento de ratificación de la Convención Marco de las Naciones Unidas sobre el Cambio Climático, hecho en Nueva York el 9 de mayo de 1992. *BOE Núm. 27*, 3125–3136. Retrieved from <https://www.boe.es/buscar/doc.php?id=BOE-A-1994-2194>
- Sun. (2010). *Hyperspectral imaging for food quality analysis and control*. Elsevier.
- Taub, D. R. (2010). Effects of Rising Atmospheric Concentrations of Carbon Dioxide on Plants. Retrieved October 14, 2021, from <https://www.nature.com/scitable/knowledge/library/effects-of-rising-atmospheric-concentrations-of-carbon-13254108/#>
- Taub, D. R., & Wang, X. (2008). Why are Nitrogen Concentrations in Plant Tissues Lower under Elevated CO<sub>2</sub>? A Critical Examination of the Hypotheses. *Journal of Integrative Plant Biology*, *50*(11), 1365–1374. <https://doi.org/https://doi.org/10.1111/j.1744-7909.2008.00754.x>
- Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, *348*(6234), 571–573. <https://doi.org/10.1126/science.aaa4984>
- Varnero Moreno, M. T. (2011). Manual de biogás. *FAO*.
- WMO. (2020a). Global Climate in 2015–2019. *WMO- No. 1249*. Retrieved from [https://library.wmo.int/index.php?lvl=notice\\_display&id=21718#.YWfVyxpbZIX](https://library.wmo.int/index.php?lvl=notice_display&id=21718#.YWfVyxpbZIX)
- WMO. (2020b). The State of Greenhouse Gases in the Atmosphere Based on Global Observations through 2019-Greenhouse Gas Bulletin (GHG Bulletin). *WMO- No. 16*.
- WMO. (2021). June ends with exceptional heat. Retrieved October 14, 2021, from <https://public.wmo.int/en/media/news/june-ends-exceptional-heat>
- Wold, S., & Sjöström, M. (1998). Chemometrics, present and future success. *Chemometrics and Intelligent Laboratory Systems*, *44*(1), 3–14. [https://doi.org/https://doi.org/10.1016/S0169-7439\(98\)00075-6](https://doi.org/https://doi.org/10.1016/S0169-7439(98)00075-6)
- Wu, D., & Sun, D.-W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals. *Innovative Food Science & Emerging Technologies*, *19*, 1–14. <https://doi.org/https://doi.org/10.1016/j.ifset.2013.04.014>
- WWF. (2018). WILDLIFE IN A WARMING WORLD. Retrieved from [www.swim2birds.co.uk](http://www.swim2birds.co.uk)



Universidad Pública de Navarra  
Nafarroako Unibertsitate Publikoa

**ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA Y  
BIOCIENCIAS**

**NEKAZARITZAKO INGENIARITZAKO ETA BIOZIENTZIETAKO GOI  
MAILAKO ESKOLA TEKNIKOA**

*APLICACIÓN DE IMÁGENES HIPERESPECTRALES (HSI-NIR) PARA LA CLASIFICACIÓN DE HOJAS DE  
PATATA DE ACUERDO A LOS EFECTOS DE LA FERTILIZACIÓN NITROGENADA*

**ANEXO: FIGURAS Y TABLAS**

presentado por

*HELENA ZOZAYA VELA (e)k*

*aurkeztua*

MÁSTER UNIVERSITARIO EN INGENIERÍA AGRONÓMICA  
*NEKAZARI INGENIERITZA UNIBERTSITATE MASTERRA*

Marzo, 2022 / 2022, Martxoa



## ÍNDICE

1. Representación del efecto de los pretratamientos sobre los datos espectrales.....	1
2. Representación de las predicciones obtenidas en la validación cruzada (CV) del modelo PLS-DA para la clasificación de hojas en función del efecto de la variedad .....	5
3. Resultados de sensibilidad, especificidad y error de los modelos PLS-DA desarrollados.....	8

## 1. Representación del efecto de los pretratamientos sobre los datos espectrales

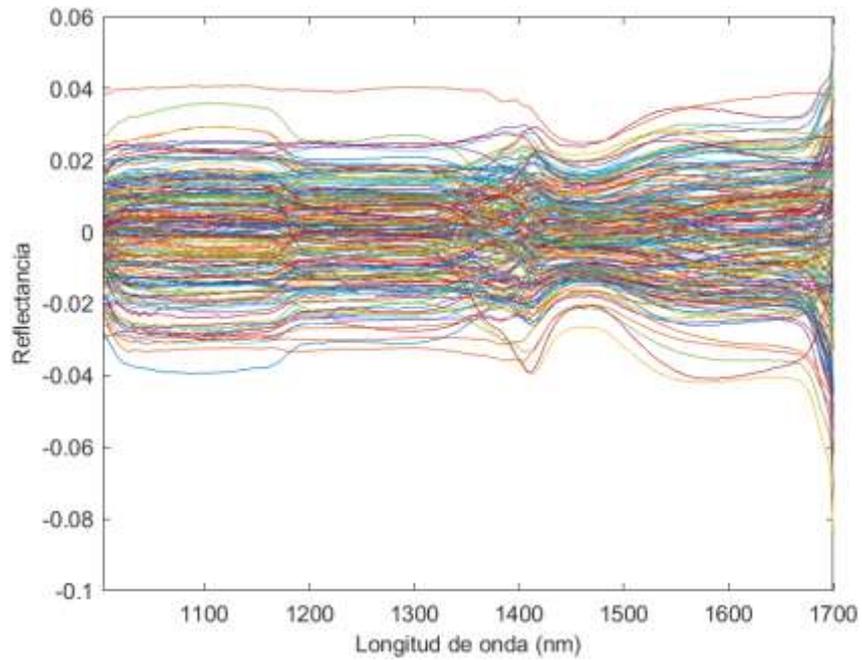


Figura 1. Efecto del pretratamiento Mean Centering en los espectros de las muestras del grupo de calibración

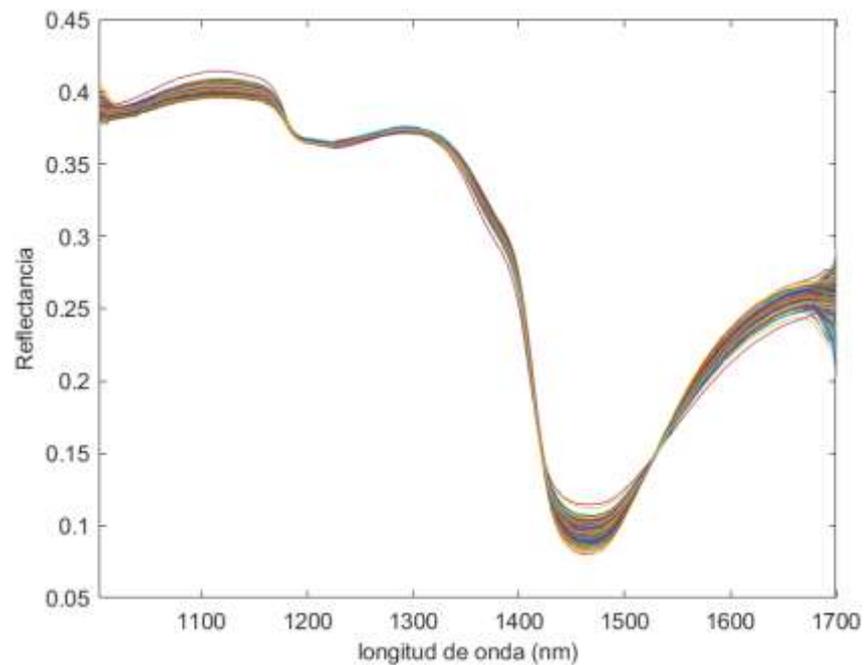


Figura 2. Efecto del pretratamiento Multiplicative Scatter Correction en los espectros de las muestras del grupo de calibración

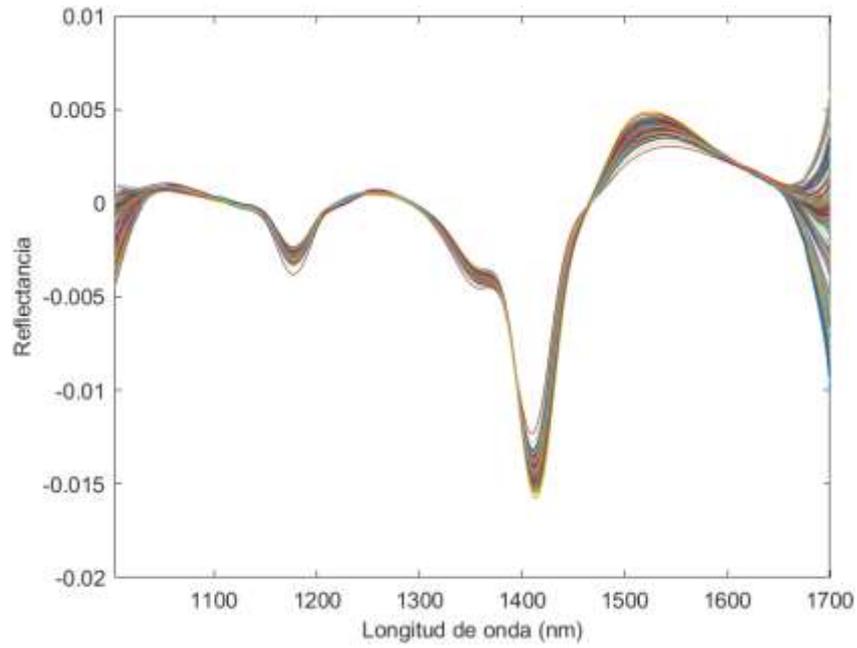


Figura 3. Efecto del pretratamiento Multiplicative Scatter Correction, en combinación con 1ª derivada, en los espectros de las muestras del grupo de calibración

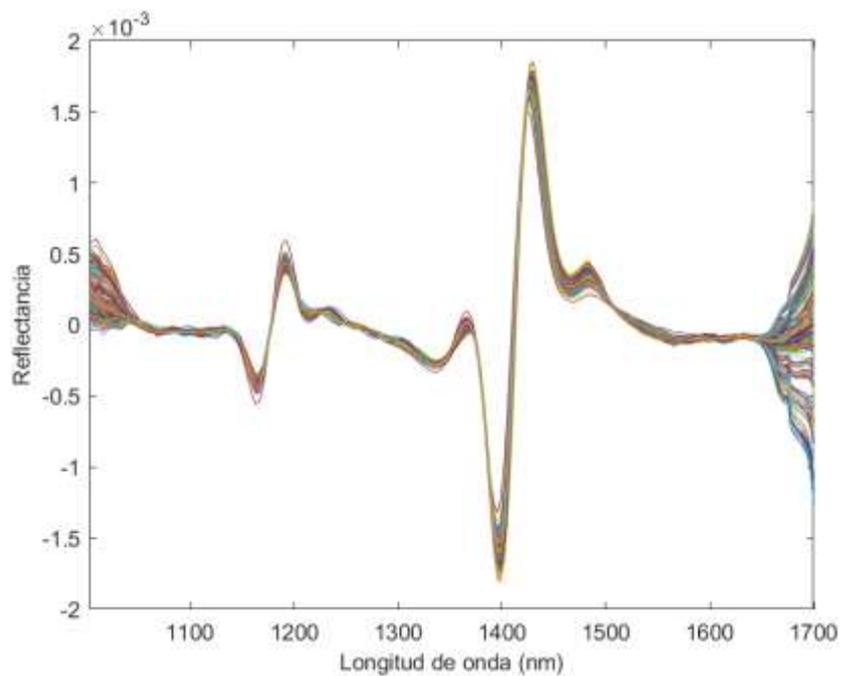


Figura 4. Efecto del pretratamiento Multiplicative Scatter Correction, en combinación con 2ª derivada, en los espectros de las muestras del grupo de calibración

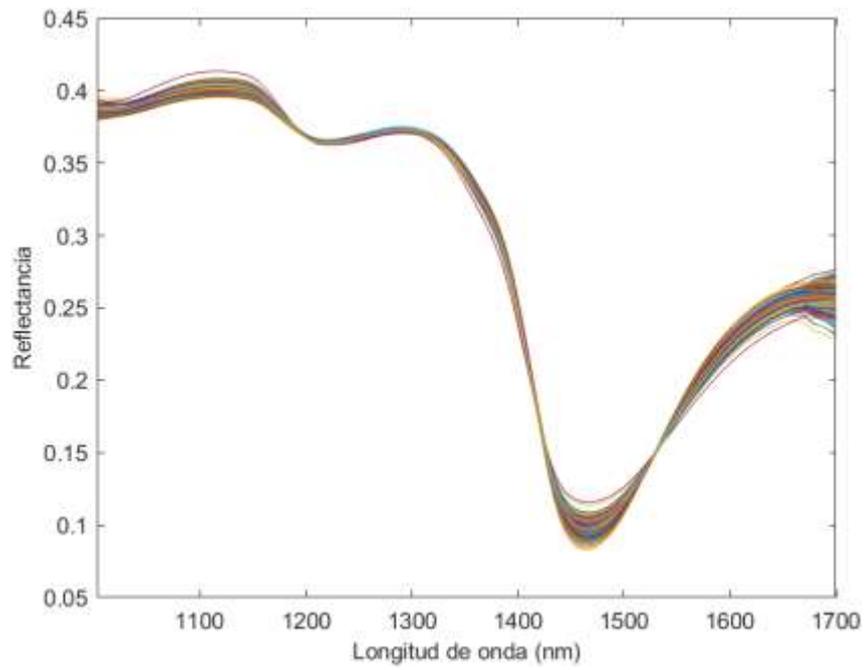


Figura 5. Efecto del pretratamiento *Multiplicative Scatter Correction*, en combinación con *Smoothing*, en los espectros de las muestras del grupo de calibración

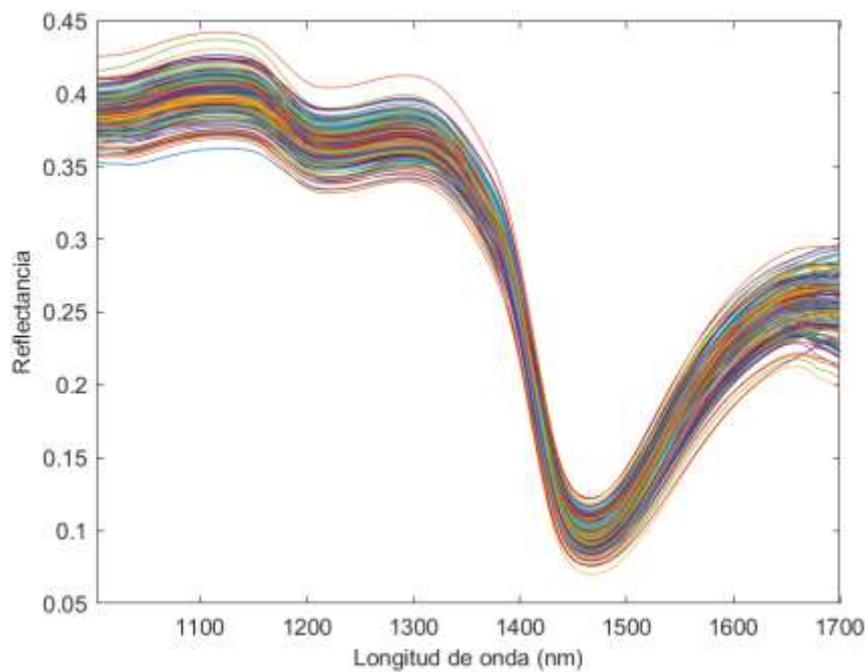


Figura 6. Efecto del pretratamiento *Smoothing* en los espectros de las muestras del grupo de calibración

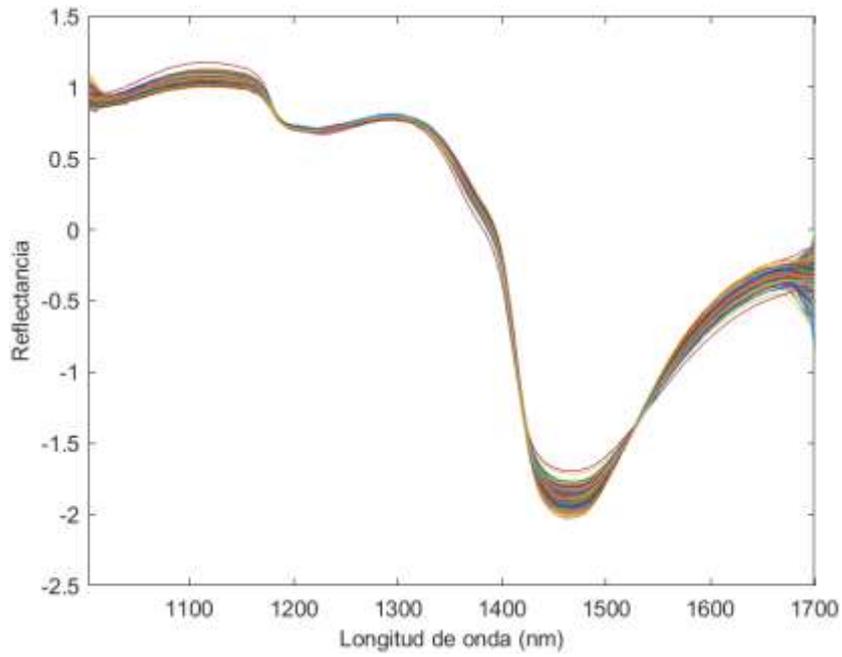


Figura 7. Efecto del pretratamiento Standard Normal Variate en los espectros de las muestras del grupo de calibración

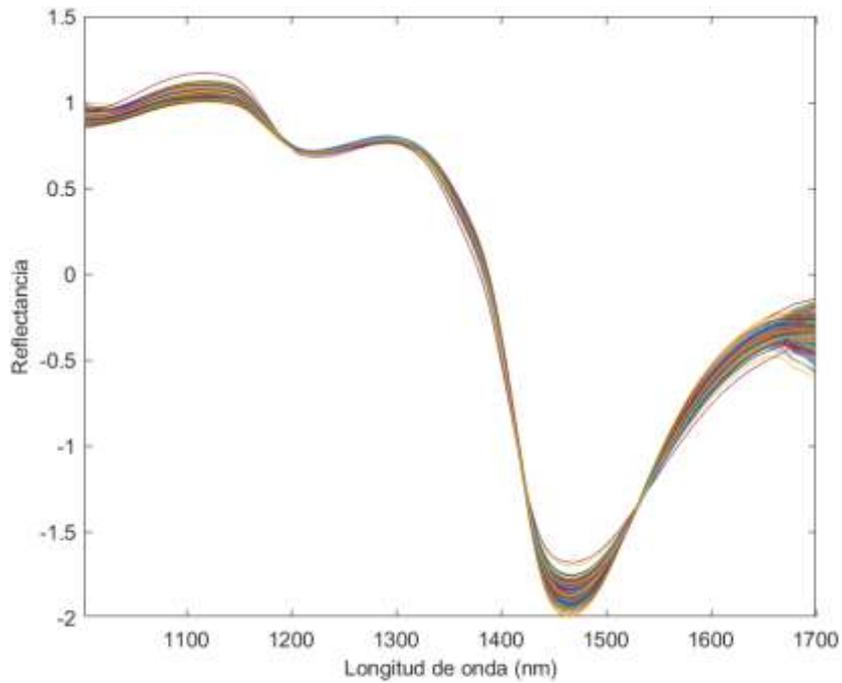


Figura 8. Efecto del pretratamiento Standard Normal Variate, en combinación con Smoothing, en los espectros de las muestras del grupo de calibración

## 2. Representación de las predicciones obtenidas en la validación cruzada (CV) del modelo PLS-DA para la clasificación de hojas en función del efecto de la variedad

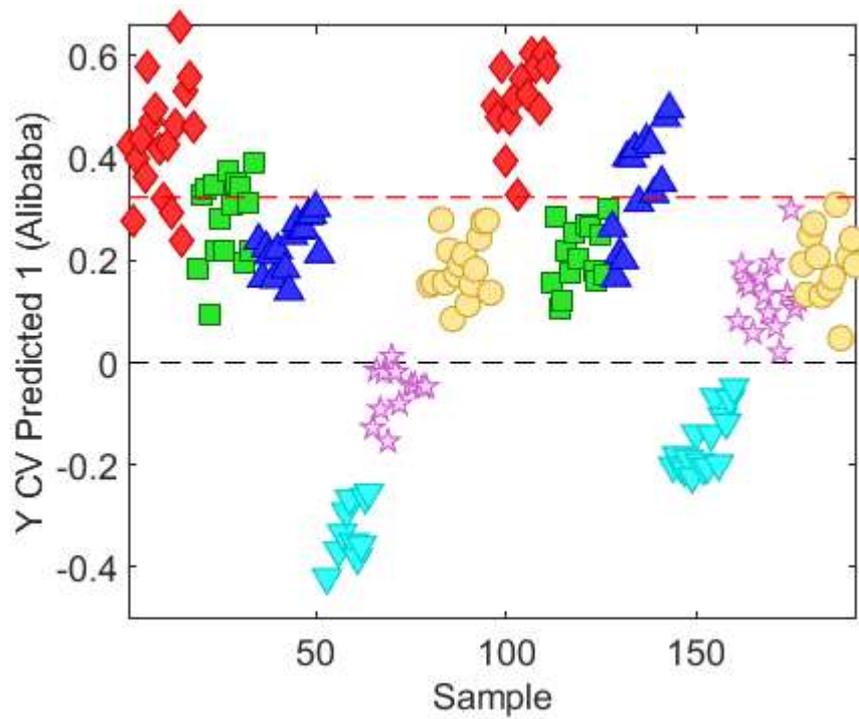


Figura 9. Predicción en la validación cruzada (CV) de la clasificación de las muestras correspondientes con la variedad Alibaba

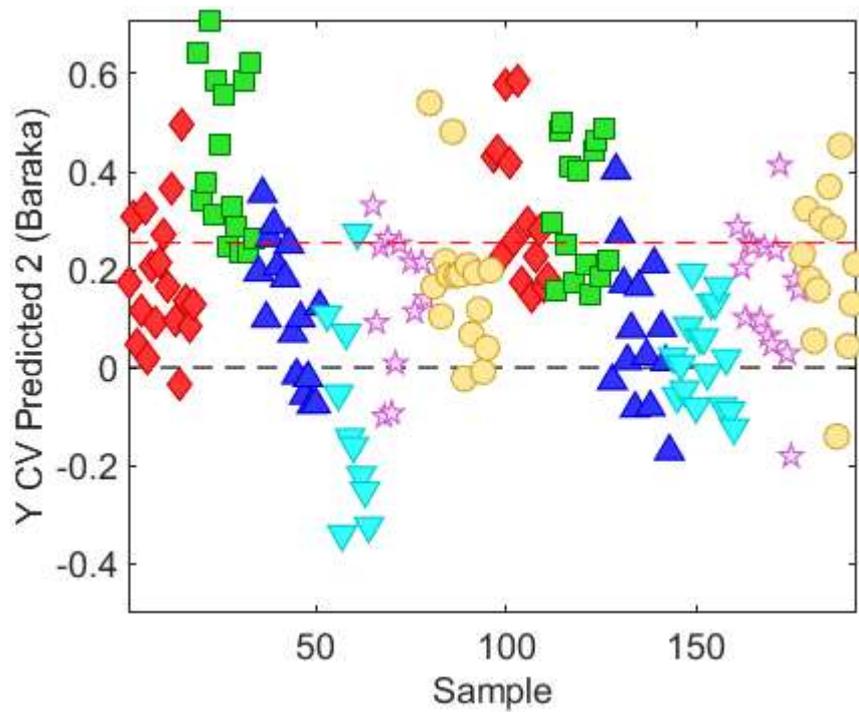


Figura 10. Predicción en la validación cruzada (CV) de la clasificación de las muestras correspondientes con la variedad Baraka

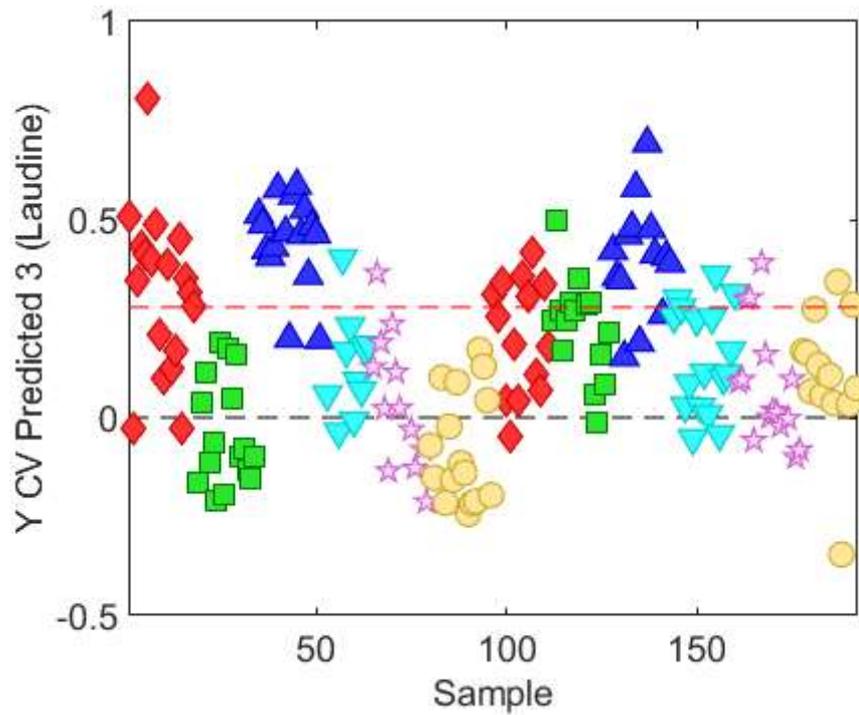


Figura 11. Predicción en la validación cruzada (CV) de la clasificación de las muestras correspondientes con la variedad Laudine

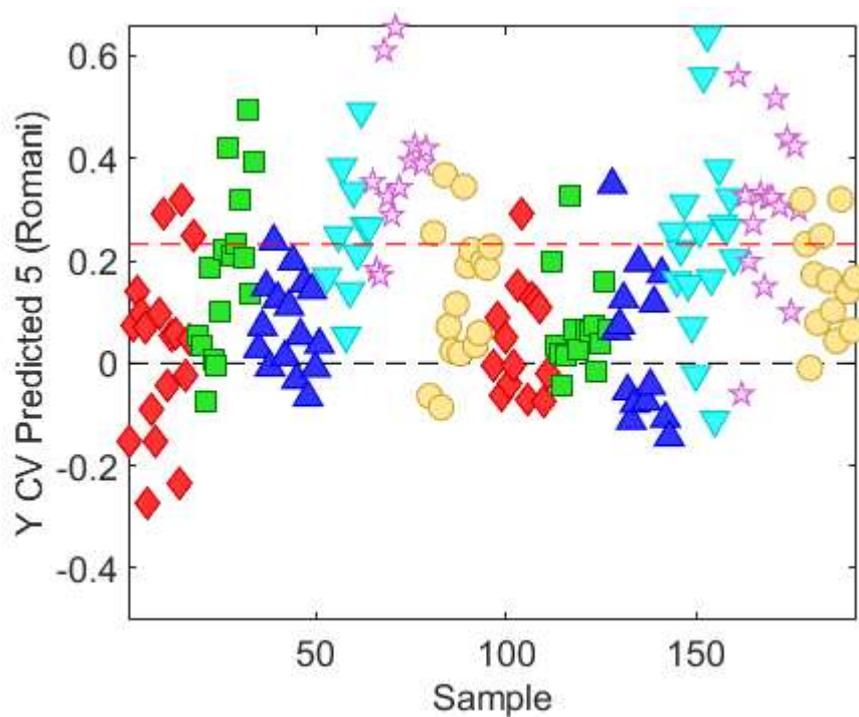


Figura 12. Predicción en la validación cruzada (CV) de la clasificación de las muestras correspondientes con la variedad Romani

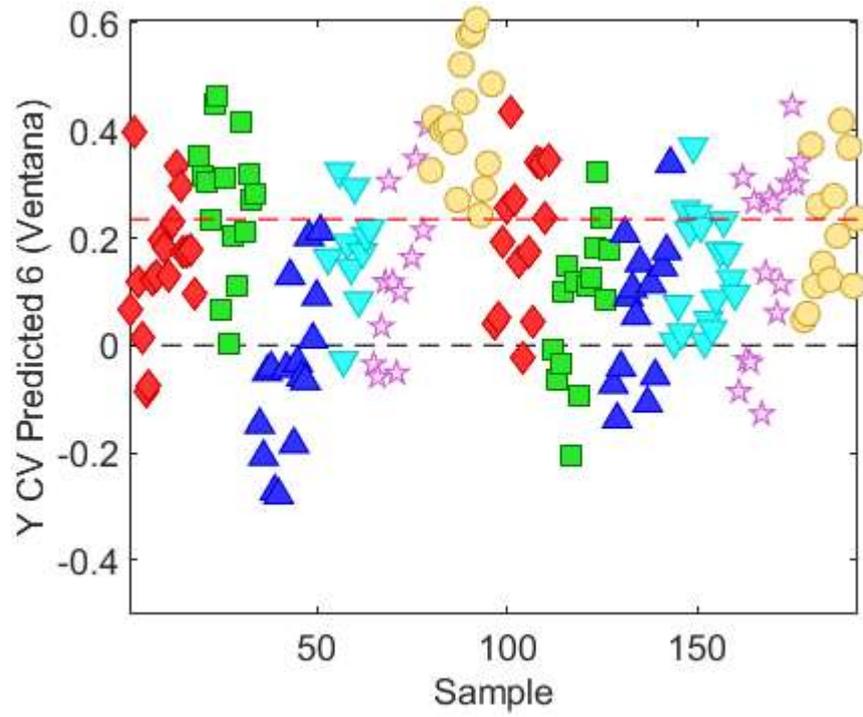


Figura 13. Predicción en la validación cruzada (CV) de la clasificación de las muestras correspondientes con la variedad Ventana

### 3. Resultados de sensibilidad, especificidad y error de los modelos PLS-DA desarrollados

Tabla 1. Resultados de sensibilidad, especificidad y error de los modelos de calibración desarrollados (Cal: Calibración; CV: Validación Cruzada; NONE: crudo; MC: Mean Centering; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; SM: Smoothing; 1ªD y 2ªD: primera y segunda derivada)

Modelo	Pretratamiento	Clase	Sensibilidad (Cal)	Especificidad (Cal)	Error (Cal)	Sensibilidad (CV)	Especificidad (CV)	Error (CV)
1	NONE	A	0,589	0,716	0,348	0,578	0,737	0,343
		B	0,716	0,589	0,348	0,737	0,578	0,343
2	MC	A	0,589	0,723	0,344	0,568	0,681	0,375
		B	0,723	0,589	0,344	0,681	0,568	0,375
3	SNV+MC	A	0,474	0,819	0,354	0,442	0,766	0,396
		B	0,819	0,474	0,354	0,766	0,442	0,396
4	MSC+MC	A	0,659	0,755	0,293	0,591	0,713	0,348
		B	0,755	0,659	0,293	0,713	0,591	0,348
5	SM+MC	A	0,484	0,853	0,332	0,432	0,800	0,384
		B	0,853	0,484	0,332	0,800	0,432	0,384
6	SNV+SM+MC	A	0,543	0,796	0,331	0,511	0,763	0,363
		B	0,796	0,543	0,331	0,763	0,511	0,363
7	MSC+SM+MC	A	0,606	0,713	0,340	0,564	0,691	0,372
		B	0,713	0,606	0,340	0,691	0,564	0,372
8	SNV+1ªD+MC	A	0,547	0,830	0,311	0,537	0,787	0,338
		B	0,830	0,547	0,311	0,787	0,537	0,338
9	SNV+2ªD+MC	A	0,628	0,763	0,304	0,638	0,710	0,326
		B	0,763	0,628	0,304	0,710	0,638	0,326
10	MSC+1ª D+MC	A	0,674	0,688	0,319	0,609	0,677	0,357
		B	0,688	0,674	0,319	0,677	0,609	0,357
11	MSC+2ªD+MC	A	0,649	0,766	0,293	0,585	0,734	0,340
		B	0,766	0,649	0,293	0,734	0,585	0,340
12	SM+MSC+MC	A	0,632	0,723	0,322	0,621	0,713	0,333
		B	0,723	0,632	0,322	0,713	0,621	0,333
13	SM+SNV+MC	A	0,685	0,688	0,314	0,663	0,699	0,319
		B	0,688	0,685	0,314	0,699	0,663	0,319