



TRABAJO DE FIN DE GRADO

**Diferenciación de grupos de edad de muestras arqueológicas mediante
la aplicación de espectrómetros NIR miniaturizados y técnicas
quimiométricas**

Fatima Zahra El Yemlahi Takkal

Supervisado por Dra. Barbara Giussani y Dr. Jordi Riu Rusell

Dpto. Química Analítica y Química Orgánica

Grado en Química

Tarragona, 2024



AGRADECIMIENTOS

Después de cuatro años de formación, estudio y dedicación continua, llega a su fin una de las etapas más significantes y gratificantes de mi vida. Durante este tiempo, he tenido el privilegio de conocer a personas que siempre ocuparán un lugar especial en mi corazón.

En primer lugar, quiero expresar mi más profundo agradecimiento a mis supervisores, la Dra. Barbara Giussani y el Dr. Jordi Riu Rusell por su guía, paciencia y conocimientos que han sido esenciales para la realización de este proyecto. También quiero agradecer al Dr. Santiago Macho Aparicio por su ayuda técnica y al Grupo Chemosens por su compañerismo.

A mi familia, gracias de todo corazón por vuestro amor, apoyo incondicional y paciencia durante este trayecto. Mamá y papá, gracias por confiar en mí y por ser mi fuente de inspiración y motivación.

Mención especial a mis compañeras de este viaje; Jessica, Alba, Siham, Somaya, Mbayang, Hafssa y Khaoula, gracias por hacer que este camino sea más ameno y por compartir tantos momentos inolvidables. Estoy segura de que nos espera un gran futuro.

Finalmente, quiero destacar a ti Lubna, por tu gran amistad y lealtad. Agradezco a la vida por haberte conocido.

¡Gracias a todos y todas!

*"El futuro pertenece a quienes creen en la
belleza de sus sueños." - Eleanor Roosevelt*

ÍNDICE

1. Resumen	4
2. Introducción	5
3. Objetivos.....	9
4. Parte experimental.....	10
4.1. Muestras.....	10
4.2. Material.....	11
4.3. Instrumentación	11
4.3.1. NeoSpectra MDK	12
4.3.2. SCiO	13
4.4. Software	14
4.5. Material.....	14
4.6. Estudio de la precisión de las réplicas	14
4.7. Análisis y procesado de datos	15
4.7.1. Análisis de componentes principales.....	15
4.7.2. Técnicas de clasificación.....	16
4.7.3. Técnicas de calibración.....	17
5. Resultados y discusión	18
5.1. Estudio de la precisión de las réplicas	18
5.2. Análisis espectral.....	19
5.3. Resultados exploratorios	20
5.4. Clasificación de las muestras por rangos de edades	24
5.4.1. Modelo de clasificación LDA	24
5.4.2. Modelo de clasificación PLS-LDA	27
5.5. Predicción de la edad en muestras arqueológicas.....	34
6. Conclusiones	37
7. Bibliografía.....	39
ANEXOS.....	42

1. Resumen

Durante los últimos años, los instrumentos NIR (Infrarrojo Cercano) miniaturizados, junto con la quimiometría, han resuelto problemas analíticos significativos en el sector de la arqueología. Estos avances han permitido realizar análisis no destructivos y en tiempo real, mejorando la eficiencia y precisión en la identificación y caracterización de materiales arqueológicos. En este trabajo, se ha evaluado la factibilidad de estos dispositivos miniaturizados para diferenciar entre muestras arqueológicas de datación reciente y aquellas con más de 100 años de antigüedad. Para ello, se han empleado dos instrumentos NIR complementarios, lo que ha permitido obtener una mayor cobertura espectral y precisión en las mediciones. Se han aplicado el PCA (Análisis de Componentes Principales) y los métodos de clasificación LDA (Análisis Discriminante Lineal) y PLS-DA (Análisis Discriminante de Mínimos Cuadrados Parciales) para el análisis cualitativo de las muestras. Estos métodos son muy eficaces para manejar datos y encontrar patrones que distinguen diferentes clases de muestras. Los resultados obtenidos demuestran la capacidad de estos instrumentos para diferenciar grupos de edad, lo que supone una herramienta valiosa para la conservación y estudio del patrimonio arqueológico.

In recent years, miniaturized NIR (Near-Infrared) instruments, combined with chemometrics, have solved significant analytical problems in the field of archaeology. These advancements have enabled non-destructive, real-time analysis, improving the efficiency and precision in the identification and characterization of archaeological materials. In this study, the feasibility of these miniaturized devices to differentiate between recently dated archaeological samples and those over 100 years old was evaluated. For this purpose, two complementary NIR instruments were used, allowing for greater spectral coverage and precision in the measurements. PCA (Principal Component Analysis), along with the LDA (Linear Discriminant Analysis) and PLS-DA (Partial Least Squares Discriminant Analysis) classification methods, were applied for the qualitative analysis of the samples. These methods are highly effective in handling large amounts of data and finding patterns that distinguish different classes of samples. The results obtained demonstrate the capability of these instruments to differentiate age groups, providing a valuable tool for the conservation and study of archaeological heritage.

2. Introducción

Diferenciar grupos de diferente antigüedad de muestras arqueológicas es importante para comprender la cronología y la evolución de los diferentes sucesos a lo largo de la historia. Sin embargo, la determinación se complica cuando el objetivo es realizar el análisis de forma instantánea y sin dañar las muestras.

Se han desarrollado técnicas para la datación de muestras arqueológicas como la datación con radiocarbono ^{14}C . El ^{14}C es un isótopo radiactivo que se forma en la atmósfera y es absorbido por los organismos vivos. Esta técnica estima la antigüedad midiendo la cantidad de ^{14}C presente en la muestra y comparándola con un estándar de referencia internacionalmente aceptado. A partir de esa cantidad, se calcula el tiempo transcurrido desde la muerte del organismo. Sin embargo, tiene algunas limitaciones, ya que es una técnica destructiva y, además, requiere muestras con un tamaño considerable, por lo que en caso de disponer de muestras pequeñas o que estén fragmentadas sería un inconveniente. Además, el proceso que se lleva a cabo suele ser costoso limitando así la accesibilidad y aplicabilidad de esta técnica.

Por ello, se han realizado nuevos estudios²⁻⁹ y se han utilizado nuevas tecnologías para superar estas limitaciones. Una de las técnicas prometedoras es la espectroscopia vibracional de infrarrojo cercano (*near-infrared*, NIR)¹⁰⁻¹². La región del infrarrojo cercano abarca un rango espectral entre aproximadamente 780 nm y 2500 nm, más allá de la luz visible, tal y como se muestra en la Figura 1.

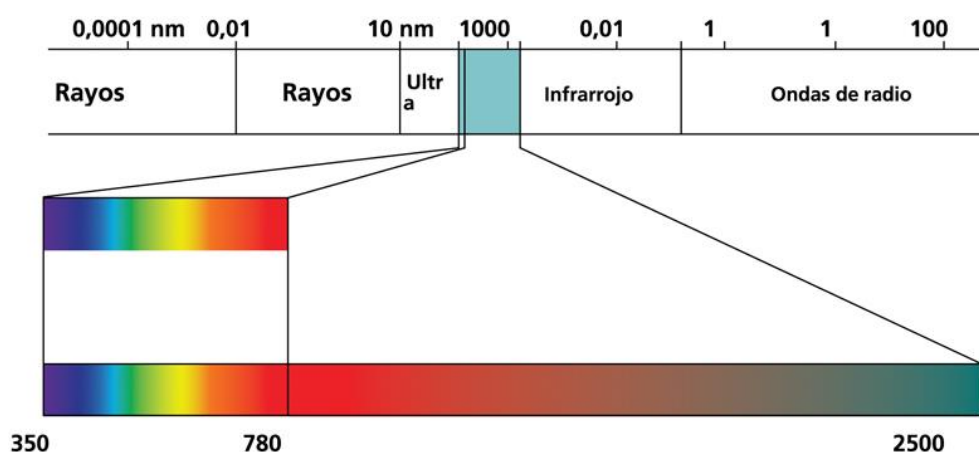


Figura 1. Espectro electromagnético con la región NIR destacada.

Los espectros NIR son gráficos que muestran la intensidad de la radiación reflejada o transmitida por una muestra. Estos surgen de las vibraciones fundamentales y sobretonos de enlaces de hidrógeno como O-H, C-H y N-H y también enlaces como C=C, C=O, C=N, lo que permite un análisis rápido de una amplia variedad de materiales con varios parámetros fisicoquímicos y composiciones químicas.

Debido al carácter multivariante de la espectroscopía NIR, es decir, al depender la información de muchas variables (longitudes de onda), normalmente se combina con métodos matemáticos y estadísticos, como la quimiometría, que es útil para extraer información de los espectros NIR. Una parte fundamental de la quimiometría es el análisis de datos multivariados, en el cual se usan técnicas para analizar datos que tienen múltiples variables de forma simultánea, lo cual es crucial tanto para los ensayos cualitativos como cuantitativos basados en el NIR.

Recientemente, se han realizado estudios^{13,14} donde se muestra la eficacia del NIR en el campo de la arqueología. La espectroscopia NIR presenta varias ventajas como su facilidad a la hora de usar. Esta permite realizar análisis *in-situ* y los pretratamientos de muestras son mínimos llegando a ser nulos. Cabe destacar que los análisis son no destructivos, en los cuales las muestras no se dañan, un factor muy importante en arqueología.

En las últimas décadas, se ha producido un avance significativo en el desarrollo de instrumentos NIR miniaturizados¹⁵⁻¹⁷. Las aplicaciones de estos instrumentos son bastante diversas, desde la agricultura y el sector alimentario hasta la ciencia de los materiales y la industria^{18,19}. Además de los beneficios que proporciona la espectroscopia NIR, existen más beneficios gracias a la portabilidad de los instrumentos miniaturizados. En primer lugar, estos instrumentos normalmente están alimentados por USB o tienen su propia batería incorporada. Suelen ser de bajo peso y pueden funcionar mediante una aplicación instalada en un *smartphone* u ordenador portátil. En segundo lugar, el bajo coste. Los espectrómetros NIR miniaturizados son más asequibles que los espectrómetros de laboratorio, permitiendo a las pequeñas y medianas empresas el acceso a este tipo de instrumentos. Finalmente, su portabilidad permite realizar análisis *in-situ* sin la necesidad de transportar la muestra hasta el laboratorio. De esta forma, se ahorra tiempo y costes de transporte²⁰⁻²³.

Actualmente, las características no destructivas, portátiles y eficientes de la espectroscopia NIR junto con la espectroscopia FTIR (Infrarrojo con Transformada de Fourier) permiten evaluar muestras arqueológicas sin dañarlas y proporcionan información detallada. En arqueología, el NIR y el FTIR se han usado en varios estudios para analizar restos esqueléticos humanos^{14,24-27}.

Por lo que respecta la espectroscopia FTIR, en un estudio²⁴ realizado el año 2017 se combina esta espectroscopia con quimiometría para estimar el intervalo post-mortem (PMI) de restos óseos humanos. Este estudio demostró que esta técnica es ideal para estimar el intervalo post-mortem. Cabe destacar también que los factores ambientales son parámetros esenciales para tener en cuenta, ya que afectan la tasa de descomposición de los restos.

Enfocando en el NIR, en un estudio¹⁴ realizado en Italia, el año 2023 se combina el NIR con la obtención de imágenes espectrales (HSI) para mapear el contenido de colágeno en huesos prehistóricos para la datación por radiocarbono. En él, se consigue revelar la distribución tanto superficial como sub-superficial de colágeno presente en la muestra.

En otro estudio¹³ realizado en Austria el año 2022, estima el PMI de restos esqueléticos humanos con espectrometría NIR portátil. Este estudio ha dado resultados positivos en los que se han podido diferenciar entre restos óseos humanos con PMI inferior a 10 años y muestras históricas de más de 100 años. No obstante, no hay estudios que revelen si los instrumentos miniaturizados son eficaces a la hora de diferenciar restos de animales. Por ello, este trabajo busca evaluar la viabilidad de utilizar sensores NIR miniaturizados para contribuir a la rápida y no invasiva determinación del rango de edad de muestras arqueológicas. Este trabajo, que se ha realizado en colaboración con el *Departament de Vertebrats del Museu de Ciències Naturals* de Barcelona, se han utilizado muestras procedentes de la colección del museo, en particular muestras de cráneos y mandíbulas de ardillas (*Sciurus vulgaris Linnaeus*).

Se han analizado dos grupos de muestras esqueléticas de ardillas, uno compuesto por esqueletos recientes, recolectados hace menos de 20 años, y otro por esqueletos

que datan de más de 100 años. Principalmente, se ha centrado en dos tipos de hueso; cráneos y mandíbulas.

Para llevar a cabo los análisis, se han optimizado las medidas para determinar si era necesario incluir réplicas instrumentales y seguidamente se ha realizado el espectro NIR. Se han usado modelos quimiométricos de exploración de datos y de clasificación para intentar encontrar diferencias entre los dos grupos de muestras que corresponden a diferentes periodos de tiempo.

Además de evaluar la viabilidad de utilizar sensores NIR miniaturizados, también se han usado dos instrumentos con características analíticas e instrumentales diferentes, SCiO y NeoSpectra MDK (Micro-Development Kit). Ambos dispositivos se usan para realizar análisis espectroscópicos en el rango del infrarrojo cercano (NIR), pero presentan diferentes capacidades analíticas y rangos de longitudes de onda. El objetivo de usar estos dos instrumentos es determinar cuál de ellos ofrece mejores resultados para resolver el problema analítico planteado.

3. Objetivos

Este trabajo tiene como objetivo realizar un estudio que permita diferenciar muestras arqueológicas con más de 100 años y muestras con menos de 20 años de antigüedad mediante la utilización de la espectroscopía NIR con instrumentos miniaturizados y técnicas quimiométricas.

Los objetivos específicos son:

- 1) Estudiar la precisión de estos instrumentos para optimizar las medidas analíticas.
- 2) Estudiar las diferencias entre 2 instrumentos NIR miniaturizados para la determinación de la edad de muestras arqueológicas.

4. Parte experimental

4.1. Muestras

Para el análisis se han usado 17 muestras de ardillas de las cuales 9 pertenecen a una datación aproximadamente de 100 años mientras que las 8 muestras restantes fueron recolectadas hace relativamente poco tiempo. La información detallada de estas muestras, incluido el lugar de recolección y el sexo, se encuentra recogida en las tablas 1 y 2.

Tabla 1. Referencias de las muestras de ardillas con datación de más de 100 años.

Referencia	Género especie	Sexo	Año recolecta
82-0080'	Sciurus vulgaris Linnaeus	Masculino	1917
82-0082'	Sciurus vulgaris Linnaeus	Femenino	1917
82-0138'	Sciurus vulgaris Linnaeus	Masculino	1920
82-0173'	Sciurus vulgaris Linnaeus	Masculino	1920
82-0174'	Sciurus vulgaris Linnaeus	Masculino	1921
82-0178'	Sciurus vulgaris Linnaeus	-	1923
82-0179'	Sciurus vulgaris Linnaeus	Femenino	1923
82-0026'	Sciurus vulgaris Linnaeus	Femenino	1923
82-0038'	Sciurus vulgaris Linnaeus	Femenino	1923

Tabla 2. Referencias de muestras de ardillas con datación reciente.

Referencia	Género especie	Sexo	Año recolecta
2006-0552'	Sciurus vulgaris Linnaeus	Masculino	2006
2010-0282'	Sciurus vulgaris Linnaeus	Masculino	2010
2010-1241'	Sciurus vulgaris Linnaeus	Masculino	2010
2017-0099'	Sciurus vulgaris Linnaeus	Femenino	2017
2017-0171'	Sciurus vulgaris Linnaeus	Masculino	2017
2018-0844'	Sciurus vulgaris Linnaeus	Masculino	2018
2019-0093-B'	Sciurus vulgaris Linnaeus	Masculino	2019
2019-0102-B'	Sciurus vulgaris Linnaeus	Masculino	2019

Las muestras se han obtenido del *Departament de Vertebrats del Museu de Ciències Naturals* de Barcelona. En la Figura 2, se pueden observar 4 de las 17 muestras.



Figura 2. Muestras de cráneos y mandíbulas de cuatro ardillas.

4.2. Material

4.3. Instrumentación

En la parte experimental de este trabajo se han utilizado dos instrumentos, el espectrofotómetro NeoSpectra MDK (Si-Ware, Cairo, Egipto) y el dispositivo SCiO (Consumer Physics, Herzliya, Israel). A continuación, se proporciona una descripción detallada de cada uno, destacando así sus características técnicas y otros datos relevantes.

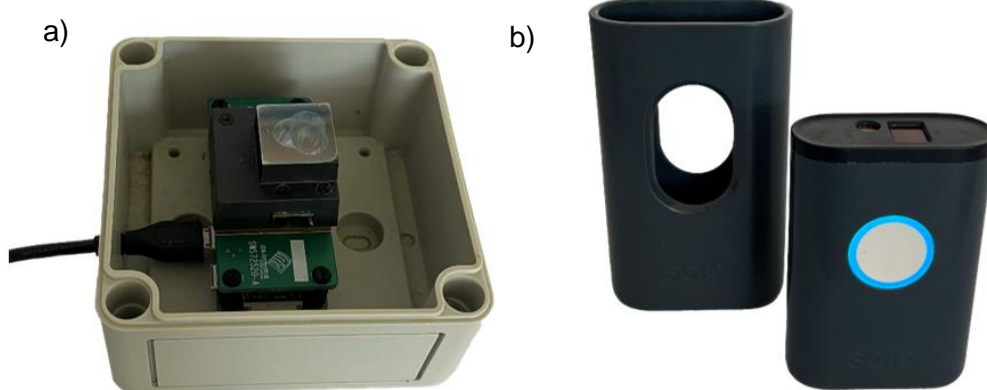


Figura 3. NeoSpectra MDK (a) y el dispositivo SCiO (b)

4.3.1. NeoSpectra MDK

El NeoSpectra Micro Development Kit (MDK) es un espectrofotómetro NIR miniaturizado desarrollado por Si-Ware, con sede en El Cairo, Egipto. Este dispositivo incluye un único fotodetector de indio, galio y arsénico (InGaAs)¹⁸.

Las dimensiones del dispositivo son 32 mm × 32 mm × 22 mm y un peso de 17 g. Su rango de longitud de onda es entre 1350 nm y 2558 nm con una resolución de 16 nm. El NeoSpectra MDK está conectado a una computadora de placa única Raspberry Pi que actúa como host y permite la conexión a través de USB a una computadora. El software permite configurar una serie de parámetros, como el tiempo de escaneo, el modo de ejecución, que puede ser simple o continuo, así como la interpolación de datos para cada espectro recopilado.

Es necesario calibrar el NeoSpectra MDK cada vez que se inicia el software. En este caso, el tiempo de escaneo se ha optimizado a 5 s. Antes de cada muestra se realiza un ajuste de fondo (*background*) utilizando el reflector SpectralonTM, un material de reflectancia del 99%. Además, antes de cada sesión, es imprescindible realizar un calentamiento del sensor durante 20 minutos para asegurar la precisión de los datos obtenidos. En la Tabla 3, se recogen las características más relevantes de este dispositivo²⁸⁻³⁰.

Tabla 3. Características técnicas del dispositivo NeoSpectra MDK

Dimensiones	32 mm × 32 mm × 22 mm
Peso	17 g
Rango espectral	1350 – 2558 nm
Resolución espectral	16 nm
Fuente de Luz	3 lámparas halógenas de tungsteno
Detector	InGaAs
Modo de trabajo	Reflectancia
Tiempo de escaneo	2 – 5 – 10 s ...
Conexión	USB

4.3.2. SCiO

El dispositivo SCiO, desarrollado por Consumer Physics en Herzliya, Israel, es un espectrofotómetro NIR miniaturizado. Este dispositivo compacto mide 67.7 mm × 40.2 mm × 18.8 mm con un peso de 35 g y puede funcionar a temperaturas que oscilan entre 4 °C y 35 °C. El dispositivo se controla a través de la aplicación "The Lab" disponible tanto para Android como para iOS y se conecta a los dispositivos móviles a través de Bluetooth. Es necesario calibrar el dispositivo cada vez que se inicia la aplicación, (utilizando el material de referencia para calibración que se encuentra en la parte posterior de la cubierta del instrumento), que suele ser un proceso rápido. Una de las ventajas del instrumento es que no requiere tiempo de calentamiento (corroborado por el fabricante) y efectivamente, los espectros obtenidos han sido estables desde el inicio del análisis¹⁸.

El SCiO opera en un rango de longitud de onda de 740 nm a 1070 nm, lo que significa que el SCiO también alcanza la parte visible del espectro electromagnético, El tiempo de adquisición varía entre de 2 y 5 segundos, con la desventaja de no poder ajustar los parámetros experimentales como sería el tiempo de exploración. Las mediciones se han realizado a una distancia de 0.5 cm respecto la muestra.

Los espectros obtenidos se almacenan en la nube y se puede acceder a ellos a través de la página web "*The Lab – Consumer Physics*". Esta función permite revisar y analizar los datos espectrales en cualquier momento y desde cualquier lugar^{29,30}.

En la Tabla 4, se mencionan las características más relevantes del SCiO.

Tabla 4. Características técnicas del dispositivo SCiO.

Dimensiones	67.7 mm × 40.2 mm × 18.8 mm
Peso	35 g
Rango espectral	740 - 1070 nm
Resolución espectral	1 nm
Fuente de Luz	Luz LED
Detector	Fotodiodo de silicio
Modo de trabajo	Reflectancia
Tiempo de escaneo	2 – 5 s
Conexión	Bluetooth

4.4. Software

En este trabajo se ha utilizado Matlab R2024a (MathWorks Inc., Natick, MA, USA) junto con el software PLS-Toolbox 9.3.1 (Eigenvector Research, Inc., Manson, WA, USA) para realizar los diferentes análisis y preprocesados. El software PLS-Toolbox, que está integrado en Matlab, está especializado en análisis multivariante.

4.5. Material

El material usado en este trabajo ha sido:

- Agua destilada
- Bastoncillos de algodón
- Guantes

Este material se ha usado principalmente para limpiar cualquier resto de polvo que haya podido quedar en las muestras.

4.6. Estudio de la precisión de las réplicas

El estudio de precisión de las réplicas se ha enfocado en analizar la variabilidad que existe entre las mediciones obtenidas a partir de réplicas analíticas (consiste en llevar a cabo múltiples mediciones en partes representativas de una misma muestra) y de réplicas instrumentales (consiste en realizar varias mediciones consecutivas, manteniendo tanto la muestra como el instrumento estables e inmovilizados).

Este estudio tiene como objetivo optimizar el tiempo de análisis, prescindiendo de las réplicas instrumentales en caso de que presenten una variabilidad insignificante en comparación con las réplicas analíticas. De esta forma, se ahorra en tiempo de medición.

Para llevar a cabo este estudio, se ha seleccionado la muestra con referencia 82-0038', a la cual se le han realizado 15 réplicas analíticas y 15 réplicas instrumentales, utilizando tanto el dispositivo NeoSpectra MDK como SCiO.

Se ha calculado la desviación estándar relativa de las 15 réplicas analíticas y de las 15 réplicas instrumentales.

4.7. Análisis y procesado de datos

Para el procesado de los datos, así como de los espectros, se han probado diferentes pretratamientos y se han seleccionado aquellos que ofrecen mejores resultados. Los preprocesados utilizados han sido el suavizado Savitzky-Golay (Smoothing Sav-Gol), la corrección multiplicativa de dispersión (MSC), la variable normal estándar (SNV), la primera derivada y la segunda derivada Savitzky-Golay (SD Sav-Gol). Varios de estos preprocesados se han empleado tanto en los modelos exploratorios como en los de regresión. Después de cada pretratamiento, y antes de calcular el modelo final, los datos se han centrado³¹.

4.7.1. Análisis de componentes principales

Para el análisis de datos exploratorio multivariante se ha empleado la técnica de Análisis de Componentes Principales (PCA), una técnica no supervisada que permite reducir la dimensionalidad del conjunto de datos manteniendo la máxima información contenida en ellos para encontrar relaciones entre los objetos y variables. En lugar de las variables originales, se obtienen nuevas variables nombradas componentes principales (PC), que son la combinación lineal de las variables originales y conservan la mayor parte de la información. Los componentes principales (PCs) son independientes entre sí y el primero contiene la máxima cantidad de información³².

Los componentes principales nos proporcionan una matriz de *scores* y una matriz de *loadings*. Los *scores* son las proyecciones de las muestras en el nuevo sistema de coordenadas definido por los PCs. Estos permiten representar en un gráfico bidimensional lo que, en el formato original de los datos, requeriría muchas dimensiones, tantas como longitudes de onda. Estas proyecciones representan las relaciones entre muestras y permiten detectar agrupaciones.

Por otro lado, los *loadings* son los coeficientes que definen la dirección de los PCs y representan la contribución de las variables originales a cada componente principal. Los *loadings* ayudan a interpretar los PCs y a entender que variables son más relevantes en la variabilidad que se observa en los datos^{33,34}.

4.7.2. Técnicas de clasificación

Para la clasificación se han usado dos técnicas, el Análisis Discriminante Lineal (LDA), usada para identificar la combinación lineal de características que maximiza la separación entre diferentes clases y el *Partial Least Squares Discriminant Analysis* (PLS-DA) que combina las características del modelo de regresión PLS (*Partial Least Squares*) con un análisis discriminante.

Ambos son modelos que normalmente se usan tanto para la modelización predictiva y descriptiva, como para la selección discriminante de variables. Permiten definir o diferenciar las clases y a diferencia del PCA, estas dos técnicas son supervisadas. Esto quiere decir que usan un conjunto de datos con etiquetas conocidas (por ejemplo, las clases de las muestras) para maximizar la separación entre las diferentes clases en los datos.

Para la validación interna se ha empleado el método “*leave one out cross-validation*”. Este construye tanta cantidad de submodelos como número de muestras hay, en el cual cada submodelo deja una muestra fuera que sirve para validar el submodelo.

Al trabajar con técnicas de clasificación, hay dos parámetros que se usan para ver la bondad del modelo, que son la sensibilidad y la especificidad. La sensibilidad mide la capacidad que tiene una técnica al identificar casos positivos y se ve reflejada como el porcentaje de casos positivos detectados respecto el total de positivos que hay realmente. En cambio, la especificidad mide la capacidad que tiene una técnica para identificar de forma correcta los casos negativos. Esta también se expresa en porcentajes de casos negativos detectados respecto el total de negativos que hay en realidad.

$$\text{Sensibilidad} = \frac{\text{Positivo verdadero}}{\text{Positivo verdadero} + \text{Falso Negativo}} \quad (1)$$

$$\text{Especificidad} = \frac{\text{Negativo verdadero}}{\text{Negativo Verdadero} + \text{Falso Positivo}} \quad (2)$$

4.7.3. Técnicas de calibración

Los modelos de regresión se han construido utilizando el método de los mínimos cuadrados parciales (PLS) y para la validación interna se ha empleado el método “*leave one out cross-validation*”. El PLS crea el modelo con la matriz de datos X (matriz de los espectros) y la matriz Y (antigüedad de las muestras) simultáneamente, y descompone los datos en variables latentes, de una forma parecida a como el PCA descompone los datos en PCs. Para que el modelo se considere bueno debe contener el mínimo número de variables latentes (LV) y valores altos de coeficiente de determinación (R^2) de la regresión entre las muestras medidas y el valor predicho de las mismas muestras, indicando que tan bien se ajustan los datos predichos por el modelo a los valores medidos. Otros factores importantes son el *Root Mean Square Error of Cross Validation (RMSECV)* y el *Root Mean Square Error of Cross Validation (RMSEC)*.

El RMSECV mide la precisión de un modelo al predecir valores desconocidos a partir de datos de validación cruzada, mientras que el RMSEC evalúa la precisión de un modelo durante la fase de calibración. Este último no depende del método de validación cruzada.

$$RMSECV = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

5. Resultados y discusión

5.1. Estudio de la precisión de las réplicas

Los resultados de la desviación estándar relativa (RSD) de las réplicas instrumentales y analíticas obtenidas con NeoSpectra MDK y SCiO se presentan en la Figura 4.

La RSD de las réplicas instrumentales realizadas con NeoSpectra MDK ha dado un rango de 1.0-7.0%, (pero mayoritariamente inferiores al 3%), mientras que la RSD de las réplicas analíticas ha dado un rango de 4.0-12.0%.

Por lo que respecta al SCiO, la RSD de las réplicas instrumentales ha dado un rango de 0.10 – 0.25% mientras que la RSD de las réplicas analíticas ha dado un rango de 7.60 – 8.20%.

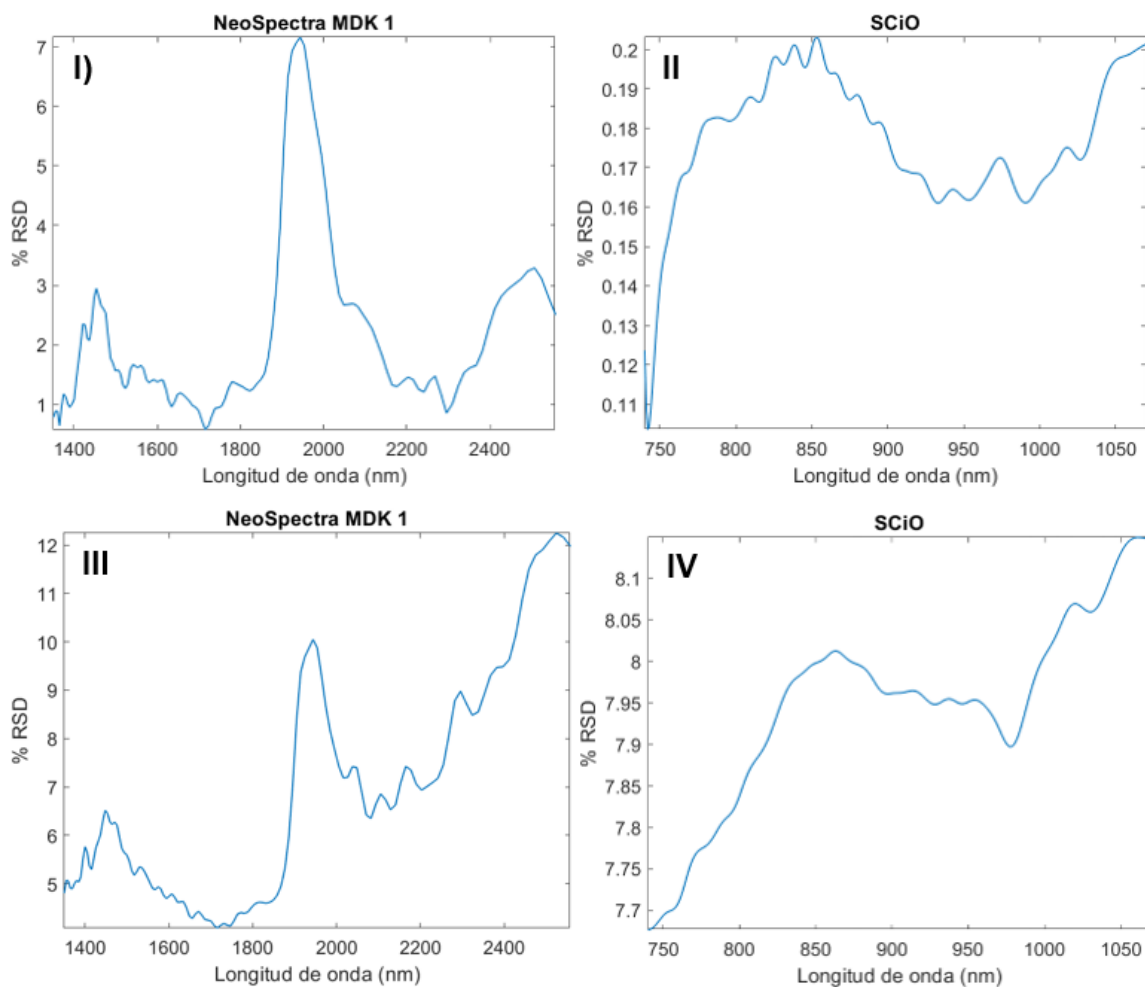


Figura 4. Resultados %RSD de las réplicas instrumentales (I – II) y analíticas (III – IV) obtenidas con NeoSpectra MDK y SCiO respectivamente.

Al analizar estos valores, se puede observar que la desviación estándar relativa (RSD) de las réplicas instrumentales es significativamente menor que la de las réplicas analíticas en ambos casos. En vista de este resultado, en el análisis de muestras de ardillas se ha decidido prescindir de las réplicas instrumentales y usar solo las analíticas.

5.2. Análisis espectral

En la Figura 5 se muestran los espectros de reflectancia que se han obtenido del análisis de las 17 muestras, tanto antiguas como recientes e incluyendo tanto cráneos como mandíbulas de cada muestra.

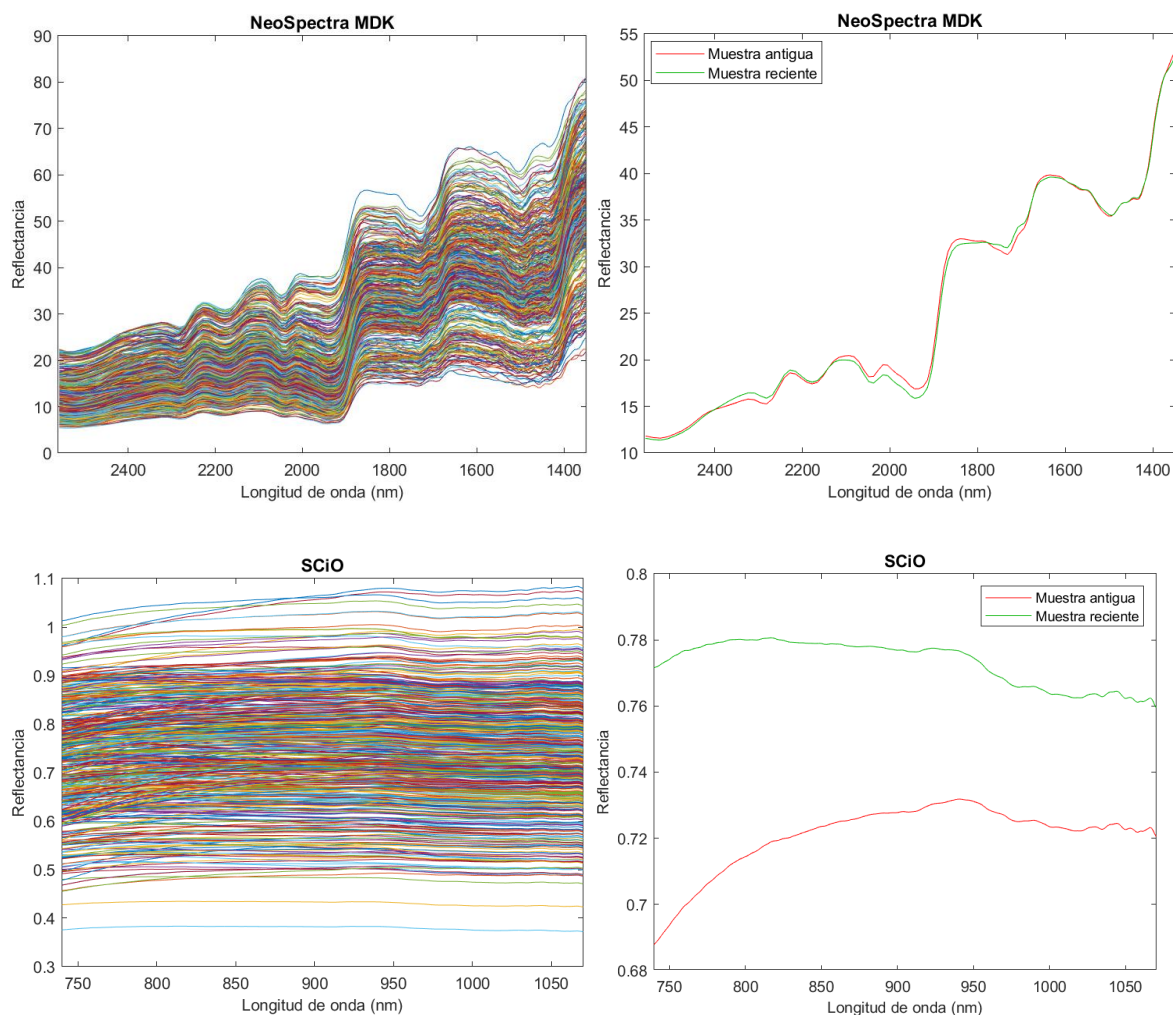


Figura 5. Espectros de muestras antiguas y recientes de cráneos y mandíbulas y del promedio de las muestras realizados con NeoSpectra MDK y SCiO respetivamente.

Los espectros se han obtenido importando los datos de los instrumentos directamente a Matlab, sin previo preprocesado.

En los espectros obtenidos con NeoSpectra MDK se pueden observar diferentes bandas y regiones. En la región entre 1670 nm y 1800 nm se indica el primer sobretono del enlace C-H. En la región entre 2000 nm y 2300 nm aparecen picos relacionados con las bandas de vibración del colágeno presente en las muestras. En más detalle, un primer pico a 2040 nm que corresponde a la combinación de los modos de estiramiento y de deformación del enlace N-H. Un segundo pico a 2120 nm que corresponde a la banda de combinación del primer sobretono del estiramiento del grupo carbonilo y la flexión del enlace N-H. Un tercer pico a 2230 nm que corresponde a la banda de combinación estiramiento y flexión del enlace de C-H^{14,35}.

En cuanto al SCiO se observa un pico a 940 nm que corresponde al tercer sobretono de vibraciones de tensión del enlace C-H.

5.3. Resultados exploratorios

Para realizar un análisis de componentes principales (PCA) se han usado los espectros promedios de las 15 réplicas analíticas de cada muestra (9 con datación de más de 100 años y 8 con datación reciente). Cabe destacar que se ha trabajado con cráneos y mandíbulas de forma separada.

En el caso de los cráneos, los datos obtenidos con NeoSpectra MDK han proporcionado una mejor separación al aplicar un suavizado Savitzky-Golay y la corrección multiplicativa de dispersión (MSC) como pretratamiento. Para las mandíbulas, se ha obtenido el valor más alto de varianza explicada en el primer PC al aplicar la segunda derivada (Sav-Gol) como pretratamiento.

En cuanto al SCiO, los cráneos se han separado mejor al aplicar un suavizado Savitzky-Golay y la variable normal estándar (SNV) como preprocesado. Para las mandíbulas, se ha obtenido una mejor separación al aplicar un suavizado Savitzky-Golay y la segunda derivada (Sav-Gol) como pretratamiento.

A continuación, se muestran los gráficos de scores obtenidos para cada instrumento.

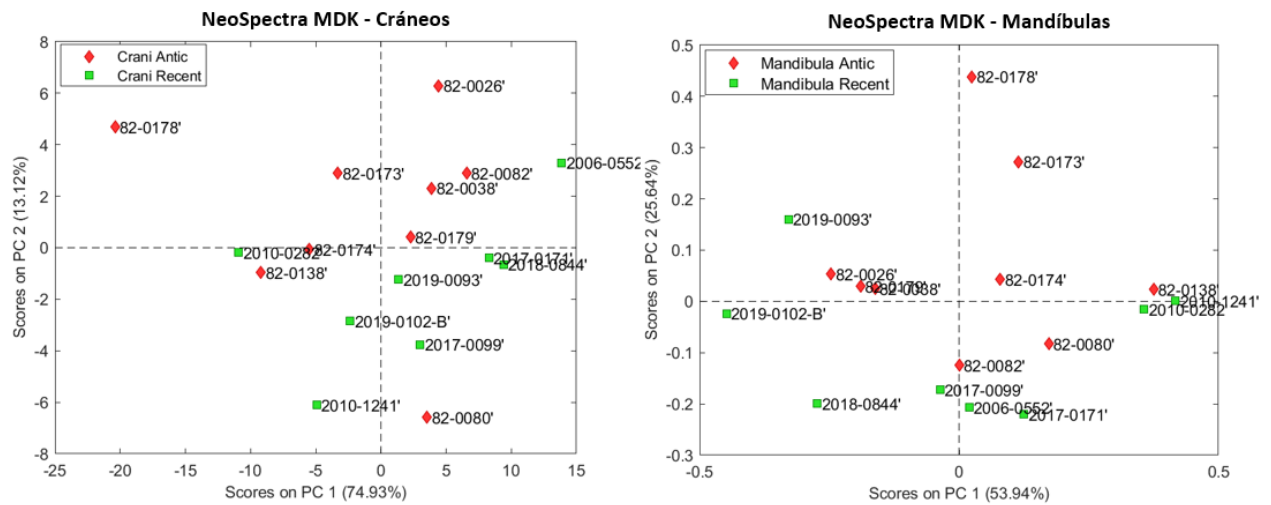


Figura 6. Gráficos de scores de NeoSpectra MDK tanto para cráneos (Suavizado, MSC) como para mandíbulas (2ª Derivada).

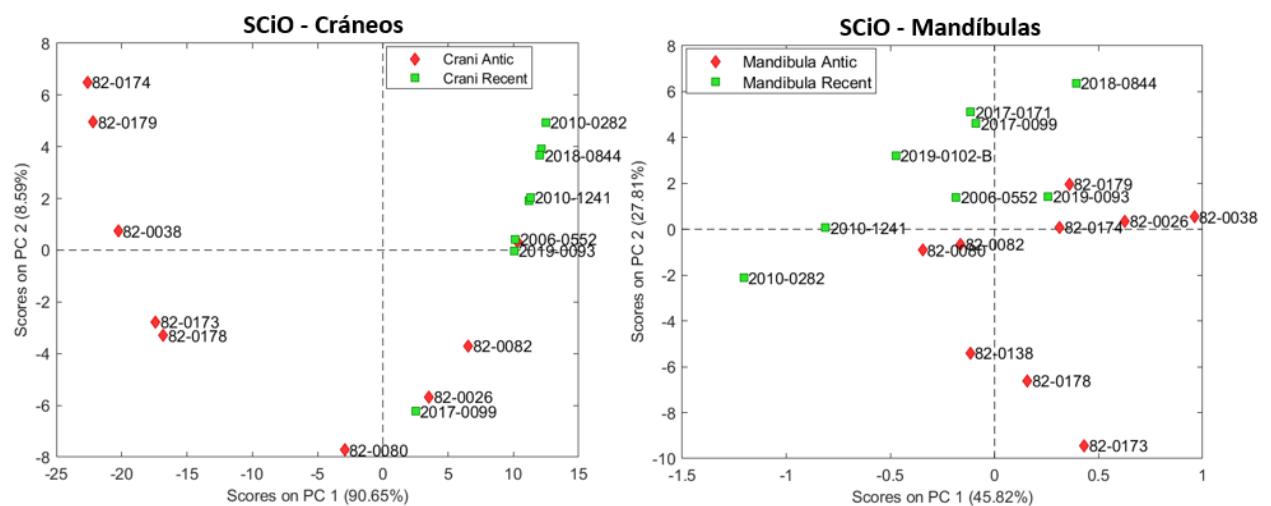


Figura 7. Gráficos de scores de SCiO tanto para cráneos (Suavizado, SNV) como para mandíbulas (Suavizado, 2ª Derivada).

Las Figuras 6 y 7 muestran el gráfico de scores para los dos primeros PCs tanto para NeoSpectra MDK como SCiO, respectivamente. En la Figura 6, se puede observar que la separación entre muestras con datación reciente (puntos verdes) y datación de más de 100 años (puntos rojos) es algo confusa e incluso se pueden considerar puntos discrepantes, como es el caso de la muestra 82-0080 o la muestra 2006-0552, posiblemente debido a la conservación.

En el caso del SCiO, la separación es considerablemente mejor tanto para cráneos como para mandíbulas. Los dos grupos de muestras, datación reciente (puntos verdes) y datación de más de 100 años (puntos rojos) se distinguen bastante bien.

La información está distribuida entre los dos primeros PCs en los cráneos con NeoSpectra MDK (70.93% PC1, 13.12% PC2) y mandíbulas (53.94% PC1, 25.64% PC2). En los cráneos con SCiO, el primer PC contiene la mayor parte de información (90.65% PC1, 8.59% PC2) mientras que las mandíbulas la información está distribuida (45.82% PC1, 27.81% PC2).

En la Figura 8, se muestran los gráficos de *loadings* correspondientes al primer y al segundo componente principal (PC) del análisis realizado con NeoSpectra MDK.

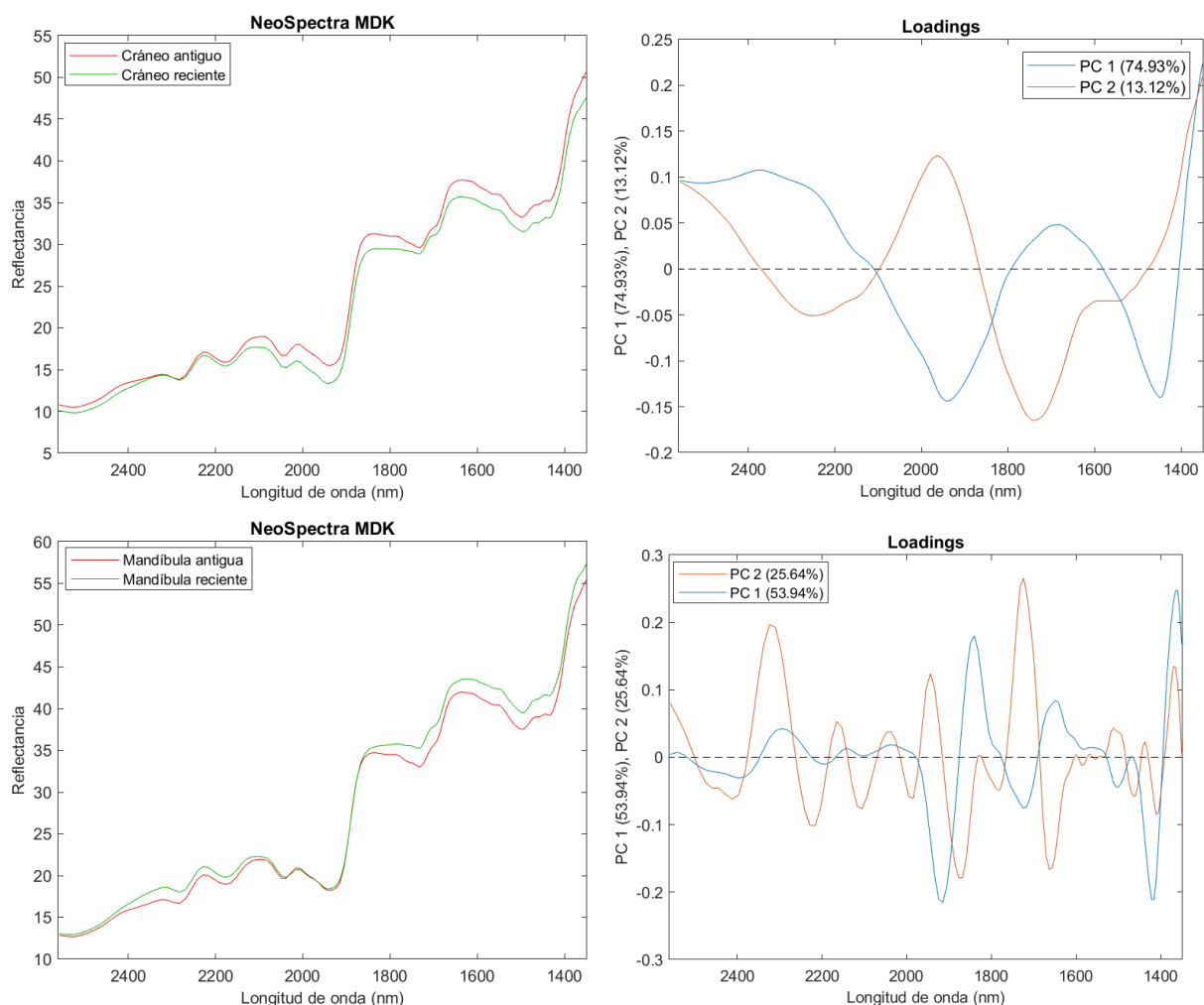


Figura 8. Promedio de las muestras realizados con NeoSpectra MDK tanto cráneos (Suavizado, MSC) como mandíbulas (2ª Derivada) y los gráficos de loadings respectivamente.

Estos gráficos ilustran las variables responsables de la separación entre las muestras de datación reciente y las muestras de más de cien años. En el caso de los cráneos, el primer PC explica la separación observada en función de las longitudes de onda que aparecen a 2230 nm y 1640 nm que coinciden con los picos presentes en el espectro y son las zonas donde los *loadings* tienen los valores más elevados en valor absoluto. El segundo PC explica la separación con las longitudes de onda que aparecen a 1950 nm y 1630 nm, que también coinciden con los picos observados en el espectro.

En el caso de las mandíbulas, el primer PC explica la separación con las longitudes de onda que aparecen a 2230 nm, 2120 nm, 2040 nm, 1952 nm y 1700 nm, que también coinciden con los picos observados en el espectro.

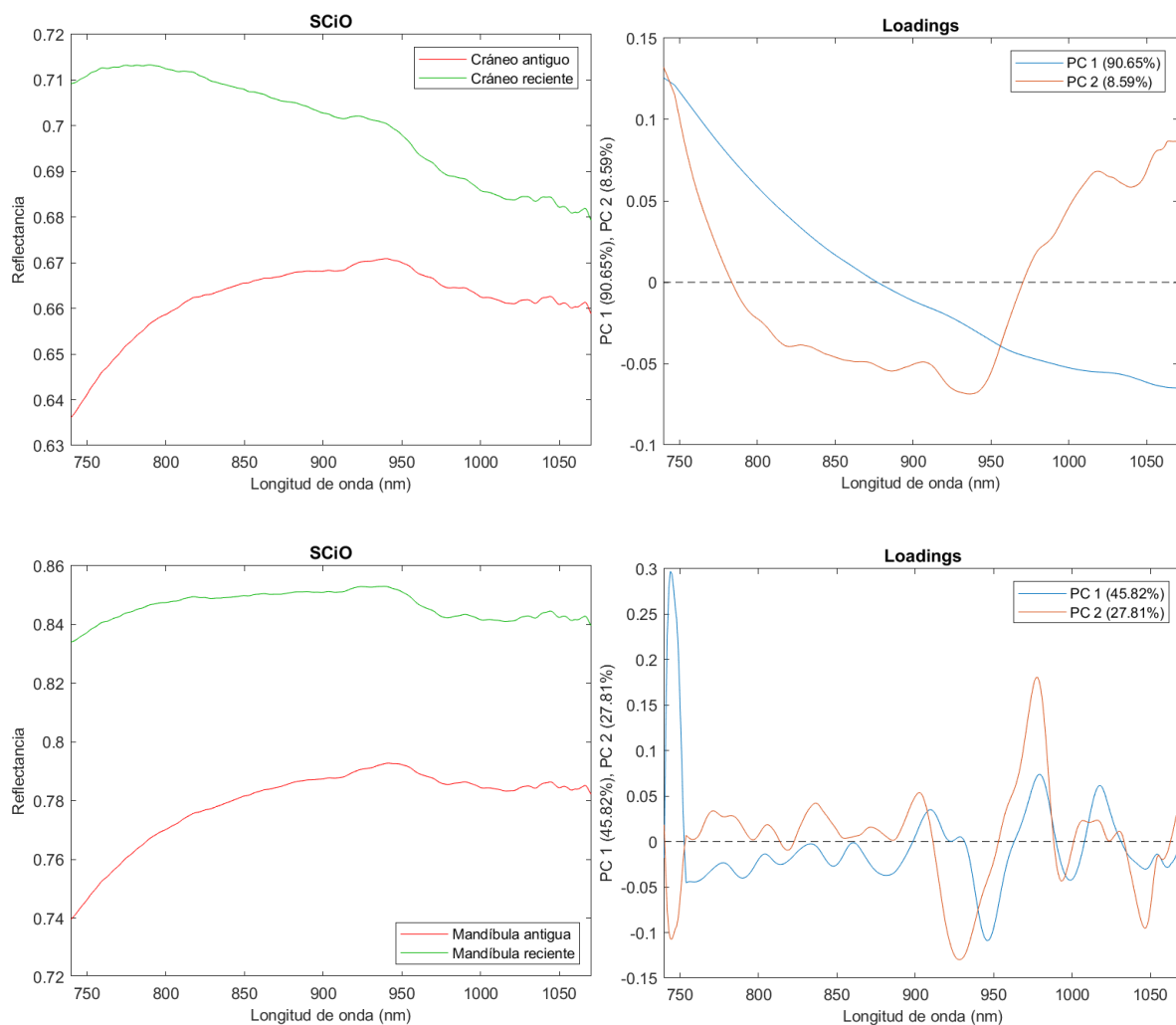


Figura 9. Promedio de las muestras realizados con SCiO tanto cráneos (Suavizado, SNV) como mandíbulas (Suavizado, 2ª Derivada) y los gráficos de loadings respectivamente.

En la Figura 9, se muestran los gráficos de *loadings* correspondientes al primer y del segundo componente principal (PC) del análisis realizado con SCiO. En el caso de los cráneos, el primer PC no explica la separación mientras que el segundo PC explica las longitudes de onda que aparecen a 940 nm y a 1020 nm que coinciden con los picos observados en el espectro.

En el caso de las mandíbulas, tanto el primer PC como el segundo, explican la separación con longitudes de onda similares, a 940 nm y a 1020 nm que coinciden con los picos del espectro.

5.4. Clasificación de las muestras por rangos de edades

5.4.1. Modelo de clasificación LDA

En este apartado se presentan y analizan los resultados obtenidos del modelo de clasificación LDA (Análisis Discriminante Lineal). El objetivo del modelo es diferenciar entre muestras arqueológicas recientes y antiguas.

El proceso del modelo comienza con la identificación de las clases que se desean distinguir dentro del conjunto de datos: muestras recientes y muestras antiguas. Una vez definidas estas clases, LDA calcula la media de cada clase y la matriz de covarianza, que mide la dispersión de los datos dentro de cada grupo.

Con las medias y la matriz de covarianza, LDA genera una función discriminante, que es una combinación lineal de las características originales (espectros en este caso). Esta función actúa como una frontera que separa las clases en un espacio de menor dimensión. Los datos originales se proyectan en este espacio definido por la función discriminante, y cada observación se clasifica en la clase más cercana a su proyección³⁶.

Se han evaluado todas las combinaciones posibles de preprocesamiento de los datos para determinar los modelos óptimos de clasificación. Los resultados incluyen valores de sensibilidad y especificidad de calibración (Cal) y de *cross validation* (CV). La diferencia entre estos dos es que la calibración utiliza el conjunto completo de los datos para ajustar el modelo y es propia del modelo mientras que el *cross validation* divide los datos en varios subconjuntos y utiliza algunos para ajustar el modelo. En

nuestro caso, nos interesa más el *cross validation* porque es una mejor estimación de la predicción de muestras futuras. Los resultados se resumen en las Figuras 10, 11, 12 y 13.

Las Figuras 10 y 11 recogen los resultados de CV relacionados con los cráneos. De los datos analizados, se puede observar que, para el dispositivo NeoSpectra MDK, los métodos de preprocesamiento como MSC, SNV, la primera derivada, la segunda derivada y la segunda derivada con suavizado (*smoothing*) son los que ofrecen los mejores resultados. Estos métodos de preprocesamiento han demostrado una sensibilidad (CV) y especificidad (CV) notablemente altas, con valores cercanos a 0.9, lo que indica que es un modelo aceptable para distinguir correctamente entre las diferentes clases analizadas.

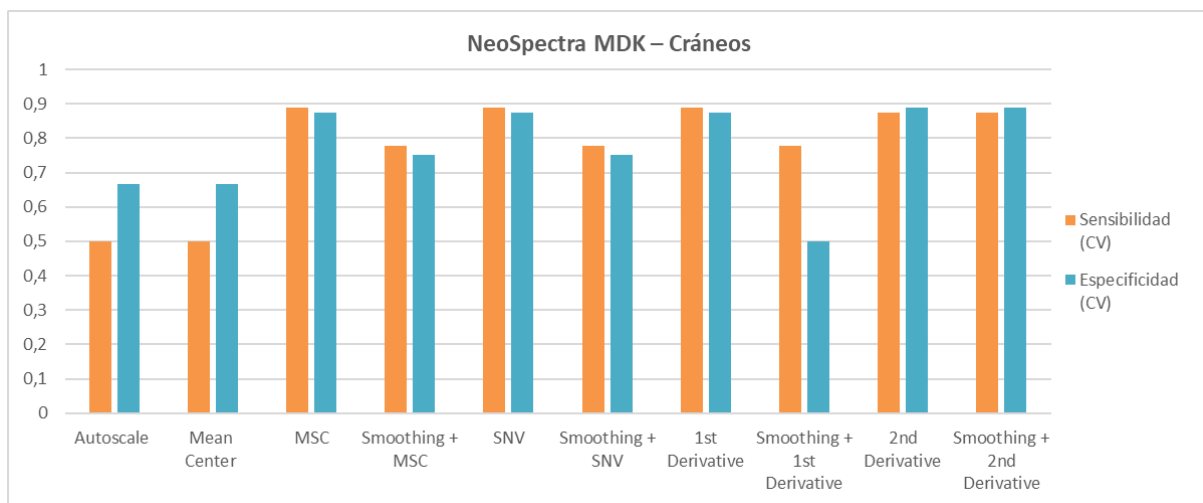


Figura 10. Resultados de los modelos de clasificación LDA de los cráneos con NeoSpectra MDK utilizando diferentes pretratamientos.

Para el dispositivo SCiO, los resultados mostrados en la Figura 11 indican que el método de preprocesamiento que ofrece el mejor resultado es la primera derivada. Con este preprocesado se han conseguido valores de 0.9 y 1 de sensibilidad (CV) y especificidad (CV) respectivamente.

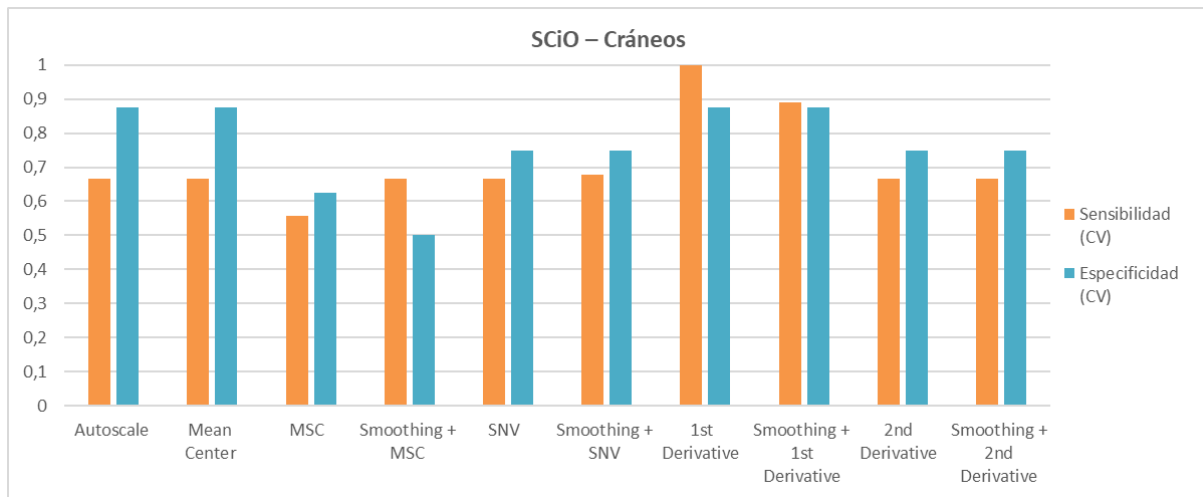


Figura 11. Resultados de los modelos de clasificación LDA de los cráneos con SCiO utilizando diferentes pretratamientos.

Las Figuras 12 y 13 presentan los resultados relacionados con las mandíbulas. En la Figura 12, que muestra los análisis realizados con NeoSpectra MDK, se observa que la mayoría de los métodos de preprocesamiento ofrecen valores cercanos a 1 en términos de especificidad y sensibilidad. En particular, el preprocesamiento combinado de suavizado y primera derivada destaca con valores de 1 en ambos factores, destacándose como el preprocesado más eficaz.

Por otro lado, en la Figura 13, los resultados para el dispositivo SCiO difieren ligeramente. Aunque no alcanzan la misma optimización que con NeoSpectra MDK, los preprocesamientos que mejor rendimiento presentan son SNV y la primera derivada. Estos métodos también muestran una alta sensibilidad y especificidad con valores de 1 y 0.875 respectivamente.

En términos generales, los análisis con mandíbulas han ofrecido valores muy altos de sensibilidad y especificidad, destacándose especialmente el pretratamiento con la **primera derivada**.

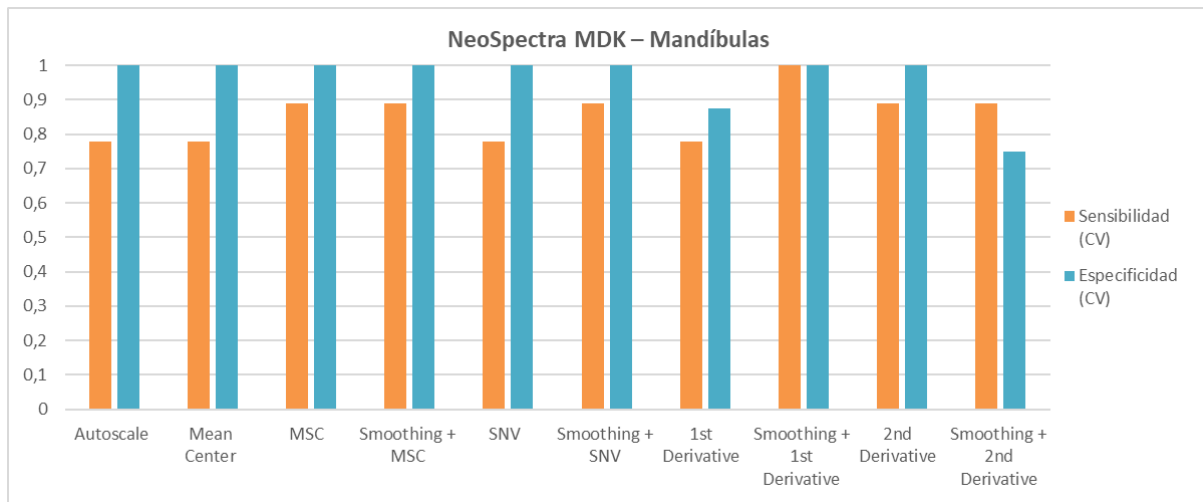


Figura 12. Resultados de los modelos de clasificación LDA de las mandíbulas con NeoSpectra MDK utilizando diferentes pretratamientos.

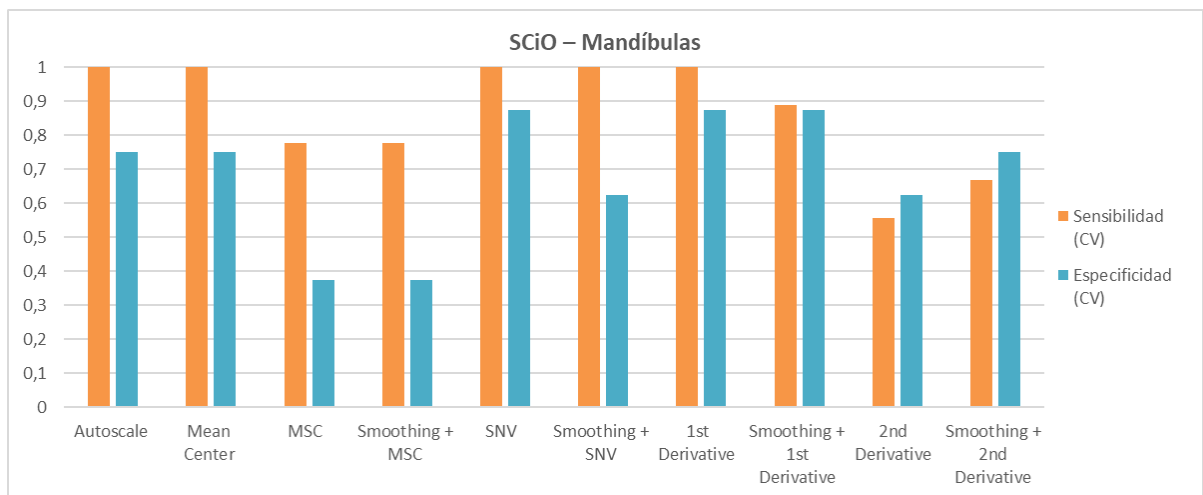


Figura 13. Resultados de los modelos de clasificación LDA de las mandíbulas con SCiO utilizando diferentes pretratamientos.

5.4.2. Modelo de clasificación PLS-LDA

A continuación, se presentan y analizan los resultados obtenidos de los mejores modelos de clasificación mediante PLS-DA (Análisis Discriminante por Mínimos Cuadrados Parciales).

En PLS-DA, primero se asignan valores numéricos a las clases a diferenciar; en este caso, se han asignado los valores de 0 y 1 de forma arbitraria, donde 0 corresponde a las muestras con una datación de más de 100 años y 1 a las muestras con datación reciente. Una vez asignadas las clases, PLS-DA construye un modelo que proyecta

las variables predictoras originales (en este caso, los espectros) en un espacio de menor dimensión, maximizando la separación entre las clases.

Este proceso de proyección se lleva a cabo mediante la descomposición de la matriz de predictores (X) y la matriz de respuesta (Y) en componentes latentes. Estos componentes latentes son seleccionados de tal manera que maximicen la covarianza entre X e Y, mientras minimizan la varianza residual en cada clase. De esta forma, el modelo PLS-DA maximiza la discriminación entre las clases³⁶.

En la Figura 14 se muestra el gráfico donde se representa la clase predicha por el modelo frente a la clase medida para los dos instrumentos. Para determinar el número óptimo de componentes, se ha seleccionado un compromiso entre el menor número posible de componentes (para evitar el sobremodelado) además de escoger buenos parámetros de clasificación (sensibilidad y especificidad). En el gráfico se puede observar una línea roja horizontal y discontinua que actúa como frontera discriminante que diferencia las predicciones entre las dos clases del modelo. Las muestras que se encuentran por encima de esta línea se clasifican en una clase, mientras que las que se encuentran por debajo se clasifican en otra.

Para este modelo se han probado diversas combinaciones de preprocesamiento y parámetros (ver [ANEXOS](#)), pero se ha seleccionado el mejor resultado para cada caso específico.

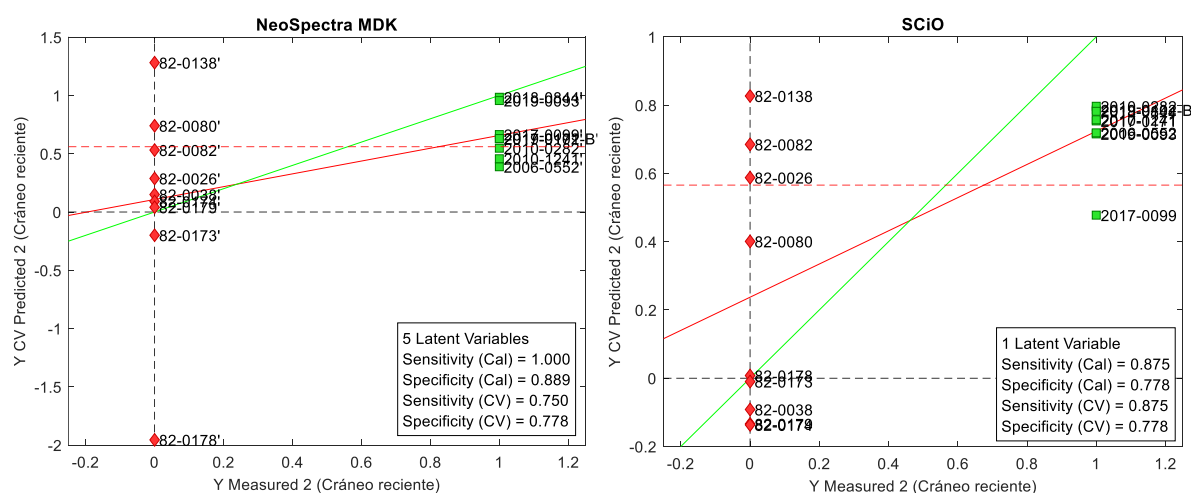


Figura 14. Gráfico entre clase predicha y clase medida de cráneos tanto para NeoSpectra MDK (Suavizado, MSC) como para SCiO (Suavizado, SNV)

Los resultados de la clasificación PLS-DA para las muestras de cráneos, de la Figura 14, demuestran un nivel de precisión aceptable. En la Tabla 5 se recogen los resultados, que incluyen los valores de sensibilidad y especificidad de calibración (Cal) y de *cross validation* (CV).

Tabla 5. Resultados de los modelos de clasificación PLS-DA de los cráneos.

Cráneos						
	Preprocesado	Nº LVs	Sensibilidad (Cal)	Especificidad (Cal)	Sensibilidad (CV)	Especificidad (CV)
NeoSpectra MDK	Suavizado MSC Centrado	5	1	0.889	0.750	0.778
SCiO	Suavizado SNV Centrado	1	0.875	0.778	0.875	0.778

Los resultados de la clasificación PLS-DA para las muestras de cráneos, de la Figura 14, demuestran un nivel de precisión aceptable.

Para el NeoSpectra MDK, es necesario emplear cinco variables latentes y, aun así, los valores de sensibilidad (CV) y especificidad (CV) no se acercan al 1, siendo 0.750 y 0.778 respectivamente. Por otro lado, el dispositivo SCiO requiere solo una variable latente para obtener resultados igualmente robustos o incluso mejores, con valores de 0.875 y 0.778.

En el caso de las mandíbulas, Figura 15, los resultados son mucho mejores en comparación con los cráneos. En la Tabla 6 se recogen los resultados.

Con NeoSpectra MDK, se necesitan cuatro variables latentes para alcanzar valores de sensibilidad y especificidad iguales a 1. Por otra parte también, con SCiO es necesario emplear 4 variables latentes para obtener valores de sensibilidad y especificidad iguales a 1. En general, los resultados son bastante buenos y permiten clasificar las muestras en dos grupos separados, destacando de nuevo como pretratamiento la **primera derivada**.

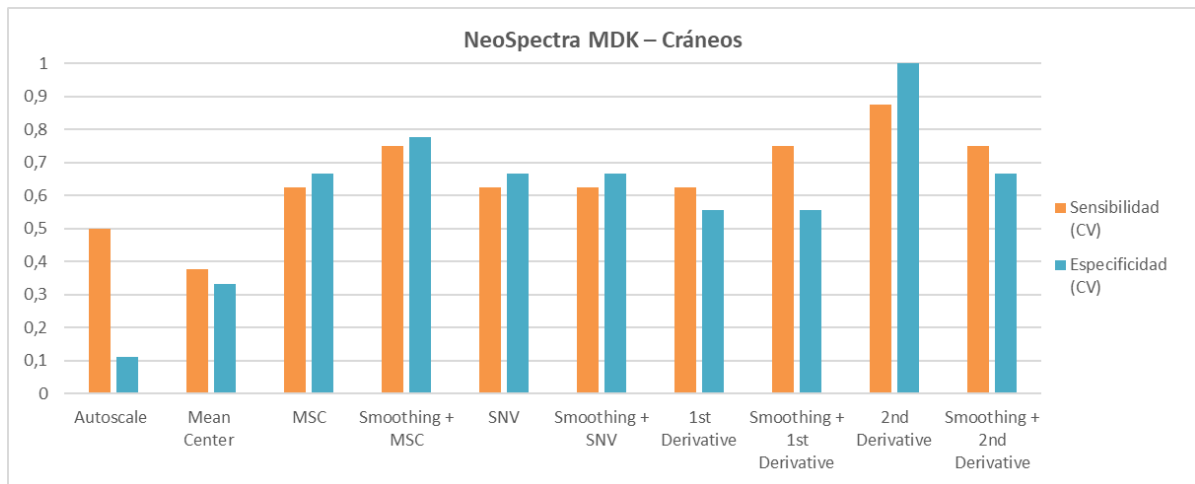


Figura 16. Resultados de los modelos de clasificación PLS-DA de los cráneos con NeoSpectra MDK utilizando diferentes pretratamientos.

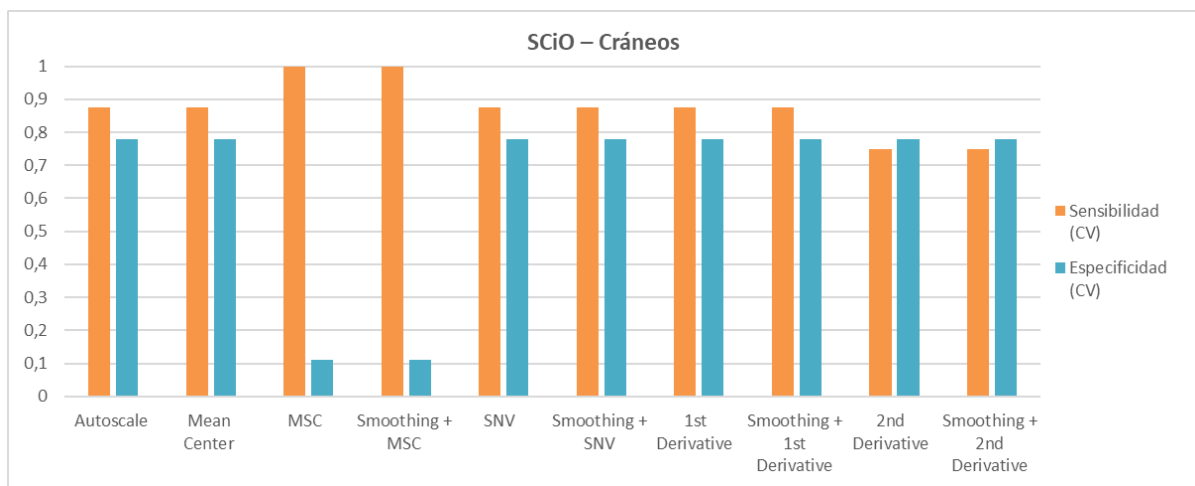


Figura 17. Resultados de los modelos de clasificación PLS-DA de los cráneos con SCiO utilizando diferentes pretratamientos.

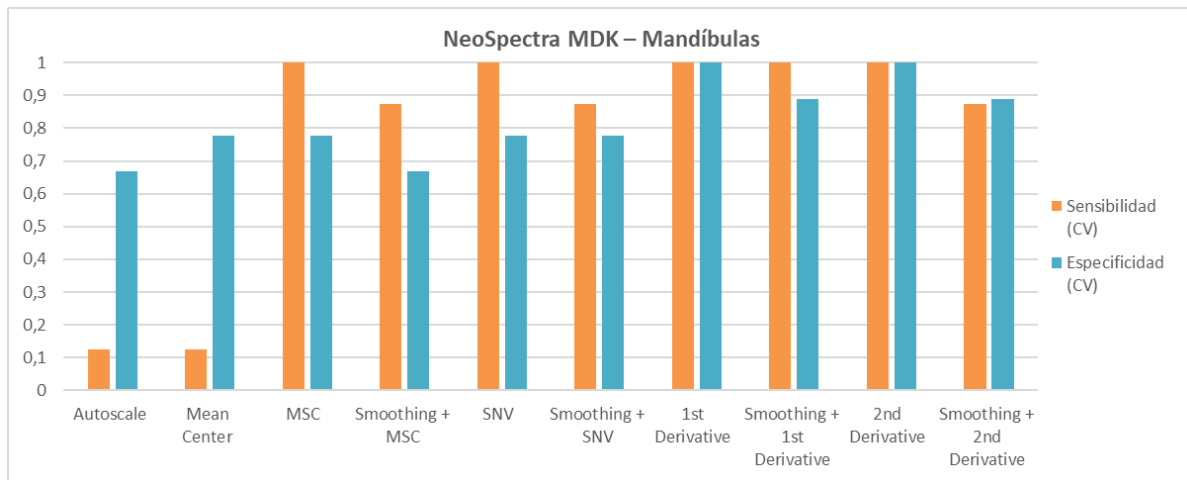


Figura 18. Resultados de los modelos de clasificación PLS-DA de las mandíbulas con NeoSpectra MDK utilizando diferentes pretratamientos.

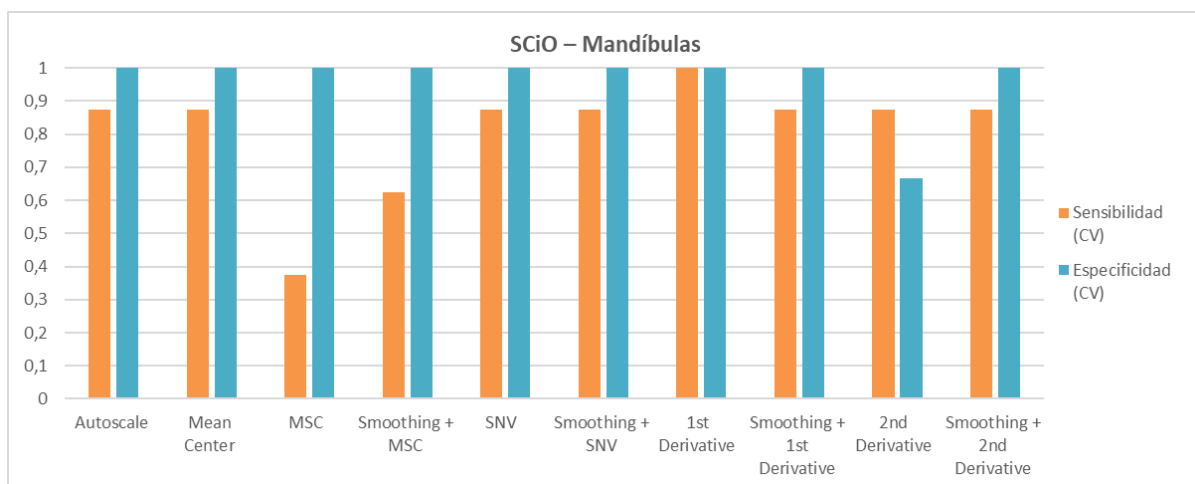


Figura 19. Resultados de los modelos de clasificación PLS-DA de las mandíbulas con SCiO utilizando diferentes pretratamientos.

En general y si se comparan los resultados de los análisis con cráneos y con mandíbulas, se puede observar claramente que con las mandíbulas se obtiene una mayor precisión en los dos instrumentos. Esto podría deberse a que las mandíbulas, al ser más pequeñas, presentan una mayor homogeneidad en su estructura, lo que podría suponer una menor variabilidad a la hora de realizar los análisis, ya que las réplicas analíticas se efectúan en puntos más cercanos entre sí. En cambio, los cráneos, al ser más grandes y complejos, pueden presentar variabilidad en su composición y estructura. Además, al tener una superficie mayor, podría implicar una mayor dificultad para obtener mediciones uniformes, lo que podría explicar la diferencia en precisión entre ambos tipos de muestras.

Al comparar los dos modelos de clasificación, LDA y PLS-DA, se observa que ambos ofrecen resultados bastante robustos. Es importante destacar que ambos modelos muestran una mayor precisión al trabajar con mandíbulas en comparación con cráneos. Sin embargo, hay algunas diferencias en el rendimiento de estos modelos según el dispositivo utilizado.

El modelo LDA tiende a funcionar mejor con el NeoSpectra MDK, mientras que el modelo PLS-DA muestra un mejor desempeño con el SCiO. A pesar de estas diferencias, cabe mencionar que la discrepancia en la precisión entre los dos modelos no es significativa, lo que sugiere que ambos modelos son complementarios para el análisis de las muestras.

5.5. Predicción de la edad en muestras arqueológicas

Se intentó aplicar el método PLS (Partial Least Squares) para correlacionar los espectros con la edad de las muestras. En este proceso, se ha desarrollado un modelo de regresión en el que la matriz X corresponde a los espectros obtenidos de las muestras, mientras que la matriz Y representa la antigüedad de cada una. En la Tabla 7 se presenta esta correlación donde cada muestra se asocia con su respectiva antigüedad.

Tabla 7. Referencias de las muestras de ardillas y su antigüedad.

Referencias antiguas	Y = Antigüedad (años)	Referencias recientes	Y = Antigüedad (años)
82-0080'	107	2006-0552'	18
82-0082'	107	2010-0282'	14
82-0138'	104	2010-1241'	14
82-0173'	104	2017-0099'	7
82-0174'	103	2017-0171'	7
82-0178'	101	2018-0844'	6
82-0179'	101	2019-0093-B'	5
82-0026'	101	2019-0102-B'	5
82-0038'	101		

Una vez asignada la antigüedad de cada muestra, se ha procedido a realizar el modelo PLS. Sin embargo, los resultados obtenidos indican que el modelo no ha sido lo suficientemente adecuado para cuantificar esta relación de manera precisa. Cabe mencionar que se han probado varias combinaciones de preprocesado y a continuación, ver Figuras 20 y 21, se muestran solo las que han ofrecido un mejor resultado.

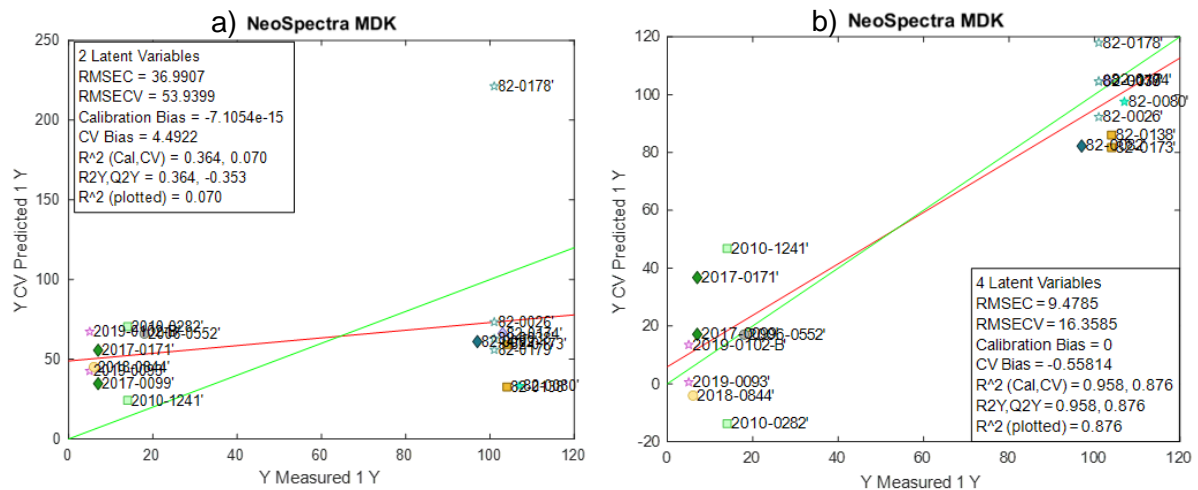


Figura 20. Resultados del modelo PLS para la determinación de la edad de muestras de ardilla – NeoSpectra MDK – Cráneos (a) y mandíbulas (b).

Para evaluar el modelo, hay que fijarse en el RMSECV. En caso de tener un RMSECV, por ejemplo, de 20, querrá decir que el error promedio que está haciendo en predecir una muestra futura es de 20 años, que se considera muy elevado.

También nos podemos fijar en el coeficiente de correlación (R^2), que es una medida estadística que indica qué tan bien se ajusta el modelo a los datos observados y su capacidad de predecir valores desconocidos que se basan en las variables de entrada. En el caso de los datos obtenidos con NeoSpectra MDK, ver Figura 20, para los cráneos se ha obtenido un valor de RMSECV de 53.94 y un coeficiente de determinación (R^2) de 0.070 al aplicar un suavizado y la primera derivada, mientras que para las mandíbulas se ha obtenido un valor de RMSECV de 16.36 y un valor de R^2 de 0.876 al aplicar la segunda derivada como preprocesado.

Por otro lado, en los datos obtenidos con SCiO, ver Figura 21, para los cráneos se ha obtenido un valor de RMSECV de 39.91 un coeficiente de determinación (R^2) de 0.312 al aplicar la segunda derivada, mientras que para las mandíbulas se ha obtenido un valor de RMSECV de 27.88 y un valor de R^2 de 0.660 al aplicar la segunda un suavizado y la segunda derivada como preprocesado.

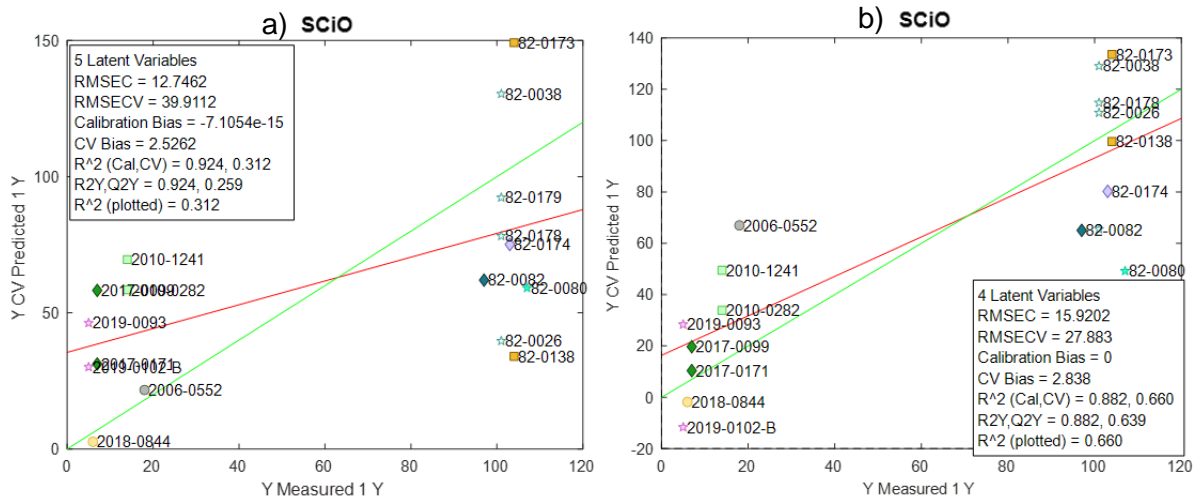


Figura 21. Resultados del modelo PLS para la determinación de la edad de muestras de ardilla – SCiO – Cráneos (a) y mandíbulas (b).

Observando los resultados, se puede ver que, los valores de RMSECV son muy elevados y los valores de R^2 inferiores a 0.9. Sin embargo, en ambos instrumentos siempre el valor de RMSECV es menor y el valor de R^2 es mayor cuando se analizan mandíbulas. Esto puede afirmar una vez más, como con los modelos de clasificación, que las mandíbulas son más eficaces para diferenciar entre los dos grupos de edad.

6. Conclusiones

Para concluir, podemos afirmar que ambos instrumentos, SCiO y NeoSpectra MDK, han demostrado ser efectivos para diferenciar entre muestras de datación reciente y de más de 100 años. Los dos dispositivos han ofrecido resultados similares, validando su utilidad en este tipo de análisis. Sin embargo, es fundamental aumentar significativamente el número de muestras para llegar a una conclusión sólida y generalizable. La aplicación de la espectroscopía NIR (Infrarrojo Cercano) tiene el potencial de ser una herramienta valiosa, pero es necesario realizar más investigaciones que incluyan una mayor variedad y cantidad de muestras.

Este trabajo representa el comienzo de una línea de investigación prometedora. Los resultados actuales sugieren que las mandíbulas ofrecen resultados más robustos a la hora de diferenciar grupos de diferente edad. No obstante, es esencial avanzar en la investigación para mejorar la precisión y fiabilidad de estos métodos.

Además, estos instrumentos miniaturizados presentan un gran interés para el sector de la arqueología, ya que son sensores económicos, no destructivos y lo suficientemente compactos como para transportarlos fácilmente. Su implementación podría transformar las prácticas de análisis, permitiendo a los profesionales efectuar análisis rápidos y precisos directamente en el lugar de trabajo.

In conclusion, we can assert that both instruments, the SCiO and the NeoSpectra MDK, have proven to be effective in differentiating between samples of recent dating and those over 100 years old. Both devices have shown similar results, validating their utility in this type of analysis. However, it is essential to significantly increase the number of samples to reach a solid and generalizable conclusion. The application of Near-Infrared (NIR) spectroscopy has the potential to be a valuable tool, but further research involving a greater variety and quantity of samples is necessary.

This work represents the beginning of a promising line of research. The current results suggest that mandibles offer more robust results when differentiating between groups of different ages. However, it is essential to advance the research to improve the precision and reliability of these methods.

Additionally, these miniaturized instruments are of great interest to the archaeology sector, as they are economical, non-destructive sensors and compact enough to be carried in a pocket. Their implementation could transform analysis practices, allowing professionals to conduct quick and accurate analyses directly in the field, thus improving real-time decision-making and reducing costs associated with traditional analysis methods.

7. Bibliografía

- (1) Gil-Chavarría, I.; Solis-Rosales, C.; Rodríguez-Ceja, M.; Chávez-Lomelí, E.; Martínez-Carrillo, M. Á.; Mondragón Sosa, M. A.; Huerta-Pacheco, N. S.; Quinto-Sanchez, M. Analysis of 14C Concentration in Teeth to Estimate the Year of Birth in the Mexican Population. *J Forensic Leg Med* **2020**, *76*, 102077. <https://doi.org/10.1016/J.JFLM.2020.102077>.
- (2) Thompson, T. J. U.; Islam, M.; Piduru, K.; Marcel, A. An Investigation into the Internal and External Variables Acting on Crystallinity Index Using Fourier Transform Infrared Spectroscopy on Unaltered and Burned Bone. *Palaeogeogr Palaeoclimatol Palaeoecol* **2011**, *299* (1–2), 168–174. <https://doi.org/10.1016/j.palaeo.2010.10.044>.
- (3) Woess, C.; Unterberger, S. H.; Roider, C.; Ritsch-Marte, M.; Pemberger, N.; Cemper-Kiesslich, J.; Hatzler-Grubwieser, P.; Parson, W.; Pallua, J. D. Assessing Various Infrared (IR) Microscopic Imaging Techniques for Post-Mortem Interval Evaluation of Human Skeletal Remains. *PLoS One* **2017**, *12* (3). <https://doi.org/10.1371/journal.pone.0174552>.
- (4) Sterzik, V.; Jung, T.; Jellinghaus, K.; Bohnert, M. Estimating the Postmortem Interval of Human Skeletal Remains by Analyzing Their Optical Behavior. *Int J Legal Med* **2016**, *130* (6), 1557–1566. <https://doi.org/10.1007/s00414-016-1395-3>.
- (5) Ortiz-Herrero, L.; Uribe, B.; Armas, L. H.; Alonso, M. L.; Sarmiento, A.; Irurita, J.; Alonso, R. M.; Maguregui, M. I.; Etxeberria, F.; Bartolomé, L. Estimation of the Post-Mortem Interval of Human Skeletal Remains Using Raman Spectroscopy and Chemometrics. *Forensic Sci Int* **2021**, *329*. <https://doi.org/10.1016/j.forsciint.2021.111087>.
- (6) Creagh, D.; Cameron, A. Estimating the Post-Mortem Interval of Skeletonized Remains: The Use of Infrared Spectroscopy and Raman Spectro-Microscopy. *Radiation Physics and Chemistry* **2017**, *137*, 225–229. <https://doi.org/10.1016/j.radphyschem.2016.03.007>.
- (7) Calabrisotto, C. S.; Fedi, M. E.; Caforio, L.; Bombardieri, L.; Mandò, P. A. Collagen Quality Indicators for Radiocarbon Dating of Bones: New Data on Bronze Age Cyprus. *Radiocarbon* **2013**, *55* (2), 472–480. <https://doi.org/10.1017/s003382220005760x>.
- (8) Biancolillo, A.; Tomassetti, M.; Bucci, R.; Izzo, S.; Candilio, F.; Marini, F. Ancient Human Bones Studied and Compared by near Infrared Spectroscopy, Thermogravimetry and Chemometrics. *J Near Infrared Spectrosc* **2019**, *27* (1), 6–14. <https://doi.org/10.1177/0967033518819417>.
- (9) Patonai, Z.; Maasz, G.; Avar, P.; Schmidt, J.; Lorand, T.; Bajnoczky, I.; Mark, L. Novel Dating Method to Distinguish between Forensic and Archeological Human Skeletal Remains by Bone Mineralization Indexes. *Int J Legal Med* **2013**, *127* (2), 529–533. <https://doi.org/10.1007/s00414-012-0785-4>.

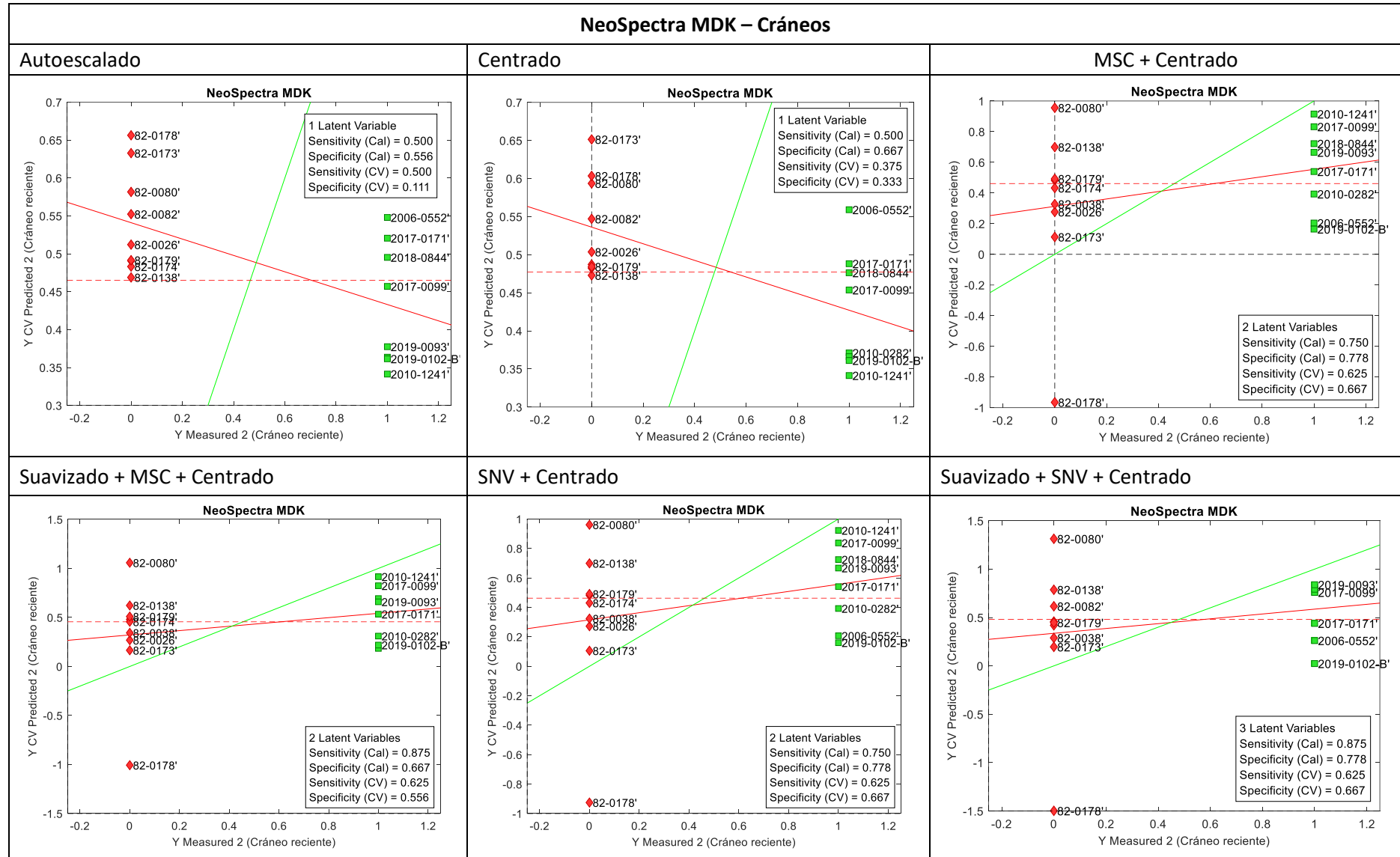
- (10) Ozaki, Y.; Huck, C.; Tsuchikawa, S.; Engelsen, S. B. Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications. In *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications*; Springer Singapore, 2020; pp 1–593. <https://doi.org/10.1007/978-981-15-8648-4>.
- (11) Dufour, É. Principles of Infrared Spectroscopy. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Elsevier, 2008; pp 1,3-1,27. <https://doi.org/10.1016/B978-0-12-374136-3.00001-8>.
- (12) Pasquini, C. Near Infrared Spectroscopy: A Mature Analytical Technique with New Perspectives – A Review. *Analytica Chimica Acta*. Elsevier B.V. October 5, 2018, pp 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>.
- (13) Schmidt, V. M.; Zelger, P.; Wöss, C.; Huck, C. W.; Arora, R.; Bechtel, E.; Stahl, A.; Brunner, A.; Zelger, B.; Schirmer, M.; Rabl, W.; Pallua, J. D. Post-Mortem Interval of Human Skeletal Remains Estimated with Handheld NIR Spectrometry. *Biology (Basel)* **2022**, *11* (7). <https://doi.org/10.3390/biology11071020>.
- (14) Malegori, C.; Sciutto, G.; Oliveri, P.; Prati, S.; Gatti, L.; Catelli, E.; Benazzi, S.; Cercatillo, S.; Paleček, D.; Mazzeo, R.; Talamo, S. Near-Infrared Hyperspectral Imaging to Map Collagen Content in Prehistoric Bones for Radiocarbon Dating. *Commun Chem* **2023**, *6* (1). <https://doi.org/10.1038/s42004-023-00848-y>.
- (15) Giussani, B.; Gorla, G.; Riu, J. Analytical Chemistry Strategies in the Use of Miniaturised NIR Instruments: An Overview. *Critical Reviews in Analytical Chemistry*. Taylor and Francis Ltd. 2024, pp 11–43. <https://doi.org/10.1080/10408347.2022.2047607>.
- (16) Rodriguez-Saona, L.; Aykas, D. P.; Borba, K. R.; Urtubia, A. Miniaturization of Optical Sensors and Their Potential for High-Throughput Screening of Foods. *Current Opinion in Food Science*. Elsevier Ltd February 1, 2020, pp 136–150. <https://doi.org/10.1016/j.cofs.2020.04.008>.
- (17) Giussani, B.; Escalante-Quiceno, A. T.; Boqué, R.; Riu, J. Measurement Strategies for the Classification of Edible Oils Using Low-Cost Miniaturised Portable Nir Instruments. *Foods* **2021**, *10* (11). <https://doi.org/10.3390/foods10112856>.
- (18) Beć, K. B.; Grabska, J.; Huck, C. W. Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers. *Chemistry - A European Journal*. Wiley-VCH Verlag January 21, 2021, pp 1514–1532. <https://doi.org/10.1002/chem.202002838>.
- (19) Correia, R. M.; Domingos, E.; Cáo, V. M.; Araujo, B. R. F.; Sena, S.; Pinheiro, L. U.; Fontes, A. M.; Aquino, L. F. M.; Ferreira, E. C.; Filgueiras, P. R.; Romão, W. Portable near Infrared Spectroscopy Applied to Fuel Quality Control. *Talanta* **2018**, *176*, 26–33. <https://doi.org/10.1016/j.talanta.2017.07.094>.

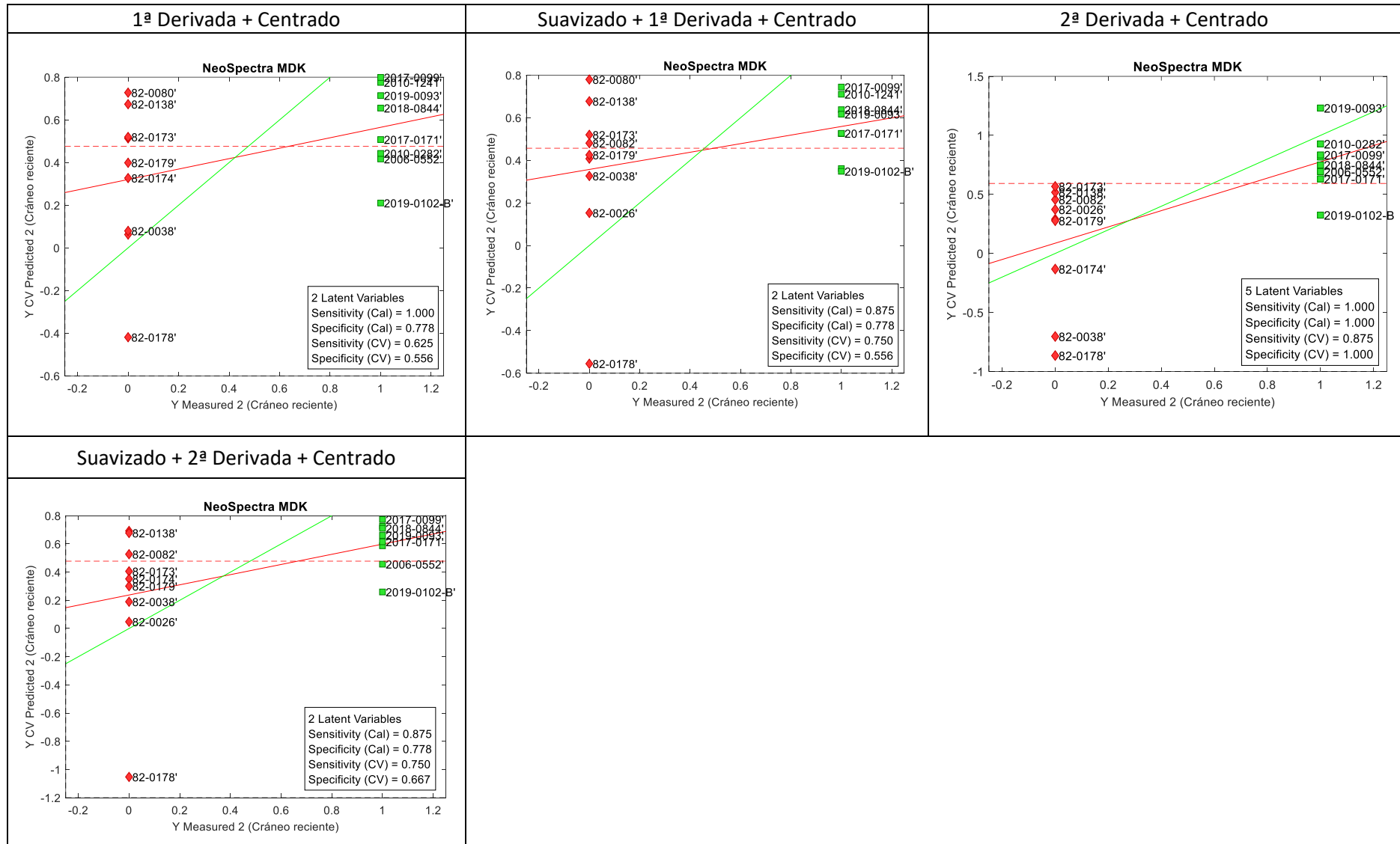
- (20) Crocombe, R. A. Portable Spectroscopy. *Appl Spectrosc* **2018**, 72 (12), 1701–1751.
<https://doi.org/10.1177/0003702818809719>.
- (21) Riu, J.; Gorla, G.; Giussani, B. Miniaturized Near-Infrared Instruments in Dairy Products or Dairy Industry: First Steps in a Long-Distance Race? *NIR news* **2021**, 32 (1–2), 17–19.
<https://doi.org/10.1177/09603360211003814>.
- (22) Beć, K. B.; Grabska, J.; Siesler, H. W.; Huck, C. W. Handheld Near-Infrared Spectrometers: Where Are We Heading? *NIR news* **2020**, 31 (3–4), 28–35.
<https://doi.org/10.1177/0960336020916815>.
- (23) Gorla, G.; Taborelli, P.; Ahmed, H. J.; Alamprese, C.; Grassi, S.; Boqué, R.; Riu, J.; Giussani, B. Miniaturized NIR Spectrometers in a Nutshell: Shining Light over Sources of Variance. *Chemosensors* **2023**, 11 (3). <https://doi.org/10.3390/chemosensors11030182>.
- (24) Wang, Q.; Zhang, Y.; Lin, H.; Zha, S.; Fang, R.; Wei, X.; Fan, S.; Wang, Z. Estimation of the Late Postmortem Interval Using FTIR Spectroscopy and Chemometrics in Human Skeletal Remains. *Forensic Sci Int* **2017**, 281, 113–120. <https://doi.org/10.1016/j.forsciint.2017.10.033>.
- (25) Longato, S.; Wöss, C.; Hatzer-Grubwieser, P.; Bauer, C.; Parson, W.; Unterberger, S. H.; Kuhn, V.; Pemberger, N.; Pallua, A. K.; Recheis, W.; Lackner, R.; Stalder, R.; Pallua, J. D. Post-Mortem Interval Estimation of Human Skeletal Remains by Micro-Computed Tomography, Mid-Infrared Microscopic Imaging and Energy Dispersive X-Ray Mapping. *Analytical Methods* **2015**, 7 (7), 2917–2927. <https://doi.org/10.1039/c4ay02943g>.
- (26) Nagy, G.; Lorand, T.; Patonai, Z.; Montsko, G.; Bajnoczky, I.; Marcsik, A.; Mark, L. Analysis of Pathological and Non-Pathological Human Skeletal Remains by FT-IR Spectroscopy. *Forensic Sci Int* **2008**, 175 (1), 55–60. <https://doi.org/10.1016/j.forsciint.2007.05.008>.
- (27) Thomas, D. B.; McGoverin, C. M.; Chinsamy, A.; Manley, M. Near Infrared Analysis of Fossil Bone from the Western Cape of South Africa. *J Near Infrared Spectrosc* **2011**, 19 (3), 151–159.
<https://doi.org/10.1255/jnirs.926>.
- (28) *NeoSpectra-Micro-SWS62231-Spectral Sensor General Description*. www.neospectra.com.
- (29) Riu, J.; Vega, A.; Boqué, R.; Giussani, B. Exploring the Analytical Complexities in Insect Powder Analysis Using Miniaturized NIR Spectroscopy. *Foods* **2022**, 11 (21).
<https://doi.org/10.3390/foods11213524>.
- (30) Riu, J.; Gorla, G.; Chakif, D.; Boqué, R.; Giussani, B. Rapid Analysis of Milk Using Low-Cost Pocket-Size NIR Spectrometers and Multivariate Analysis. *Foods* **2020**, 9 (8).
<https://doi.org/10.3390/foods9081090>.

- (31) Rinnan, Å.; Berg, F. van den; Engelsen, S. B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TrAC - Trends in Analytical Chemistry*. November 2009, pp 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>.
- (32) Bereton, R. G. Pattern Recognition. In *Applied Chemometrics for Scientists*; 2013; pp 145–191. <https://doi.org/10.1016/b978-0-444-59528-7.09993-2>.
- (33) D. L. Massart, B. G. M. V. L. M. C. B. S. D. J. P. J. L. and J. S.-V. *Handbook of Chemometrics and Qualimetrics: Part A*; 1997.
- (34) D. L. Massart, B. G. M. V. L. M. C. B. S. D. J. P. J. L. and J. S.-V. *Handbook of Chemometrics and Qualimetrics: Part B*; 1997.
- (35) Sponheimer, M.; Ryder, C. M.; Fewlass, H.; Smith, E. K.; Pestle, W. J.; Talamo, S. Saving Old Bones: A Non-Destructive Method for Bone Collagen Prescreening. *Sci Rep* **2019**, 9 (1). <https://doi.org/10.1038/s41598-019-50443-2>.
- (36) Brereton, R. G. Applied Chemometrics for Scientists. In *Applied Chemometrics for Scientists*; Wiley, 2007. <https://doi.org/10.1002/9780470057780.fmatter>.

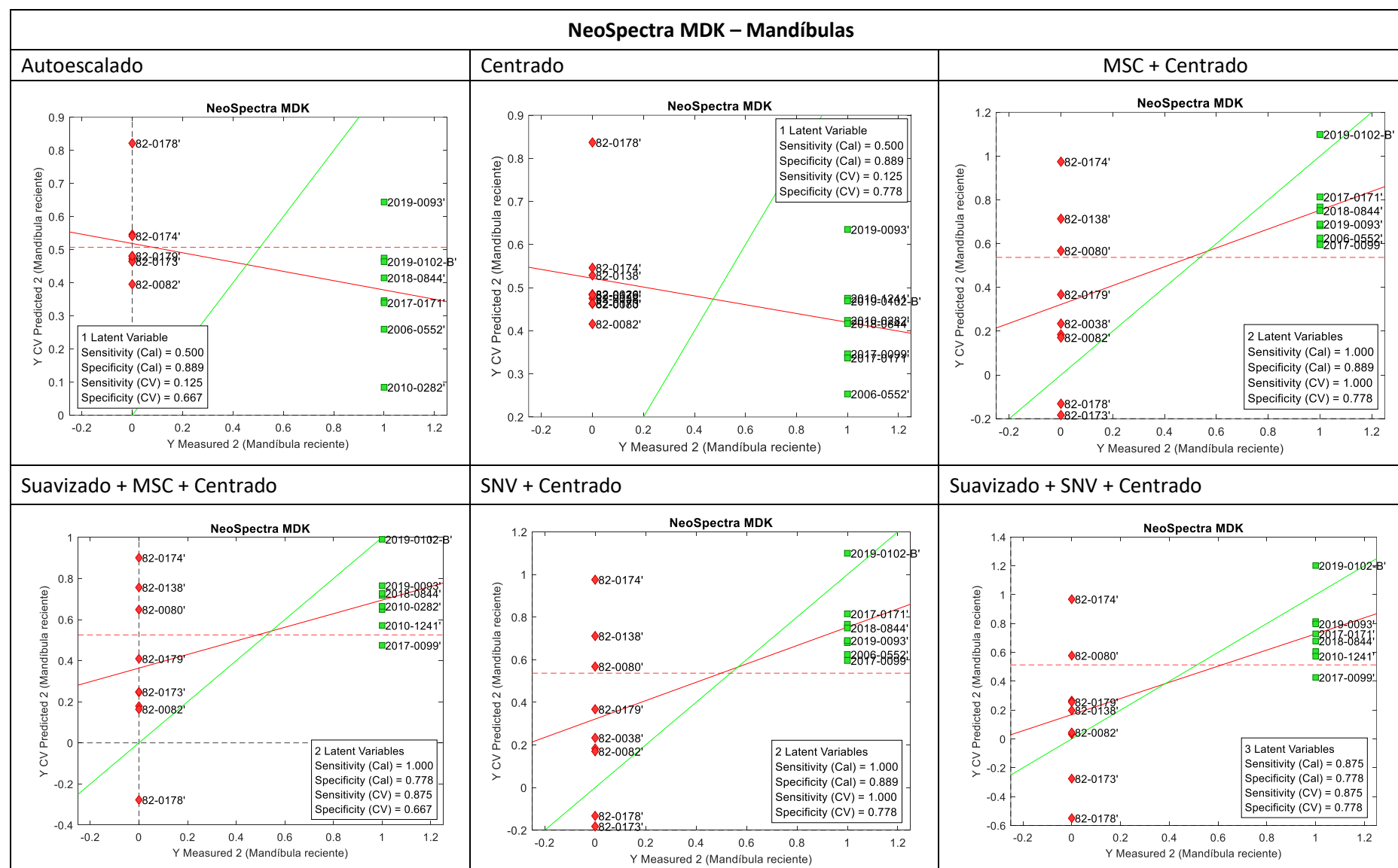
ANEXOS

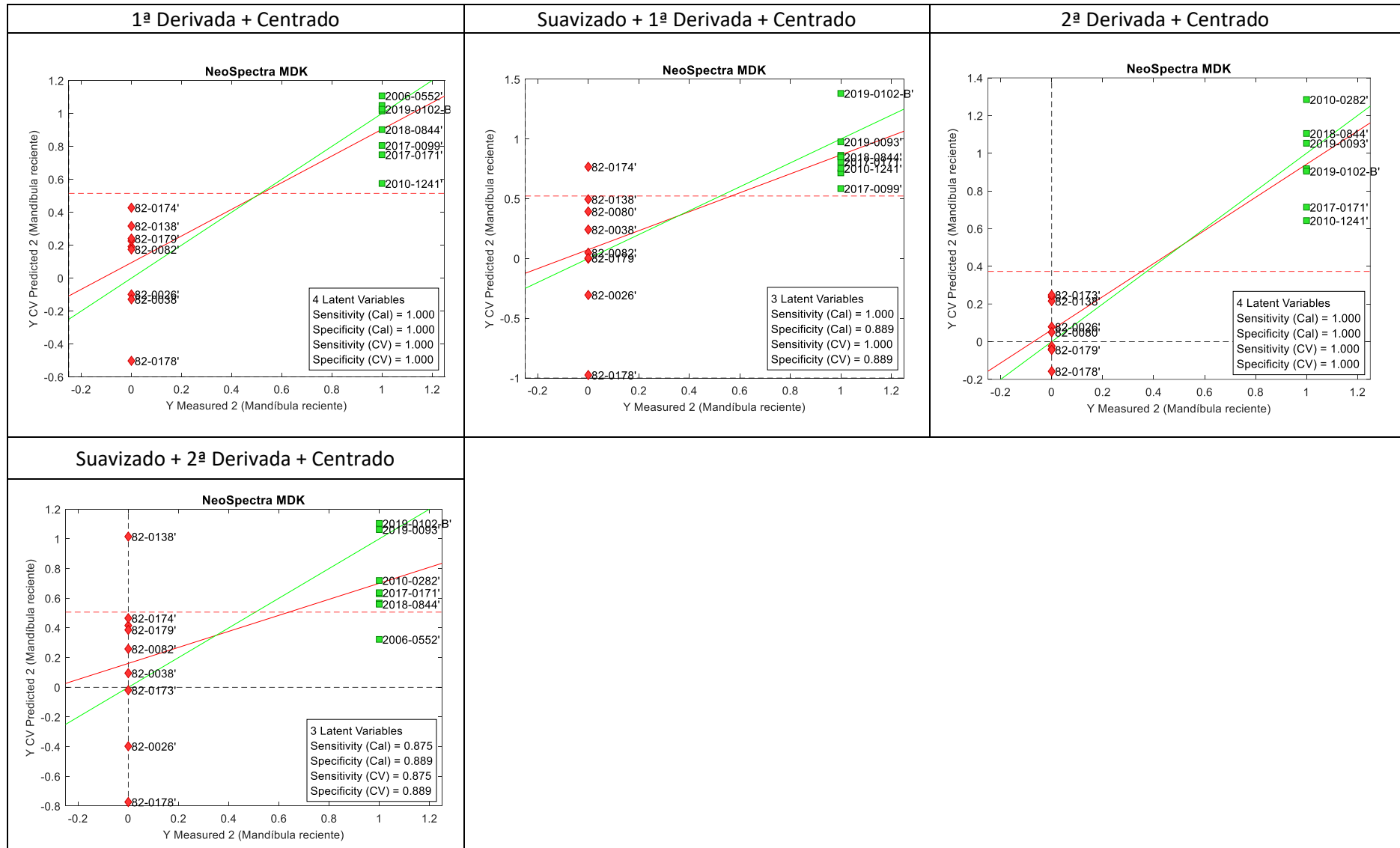
Anexo 1. Pretratamientos con el modelo de clasificación PLS-DA para cráneos analizados con NeoSpectra MDK.



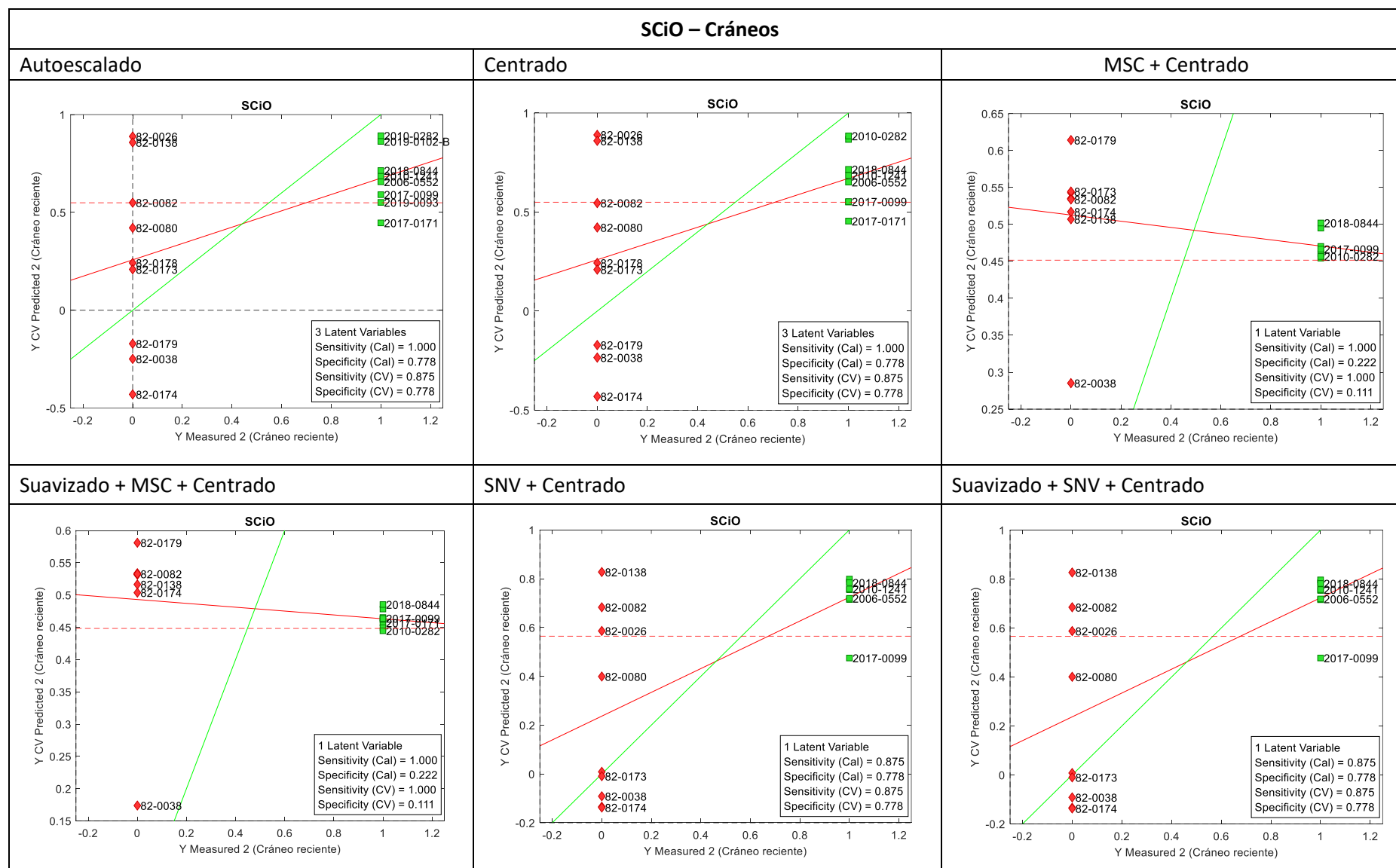


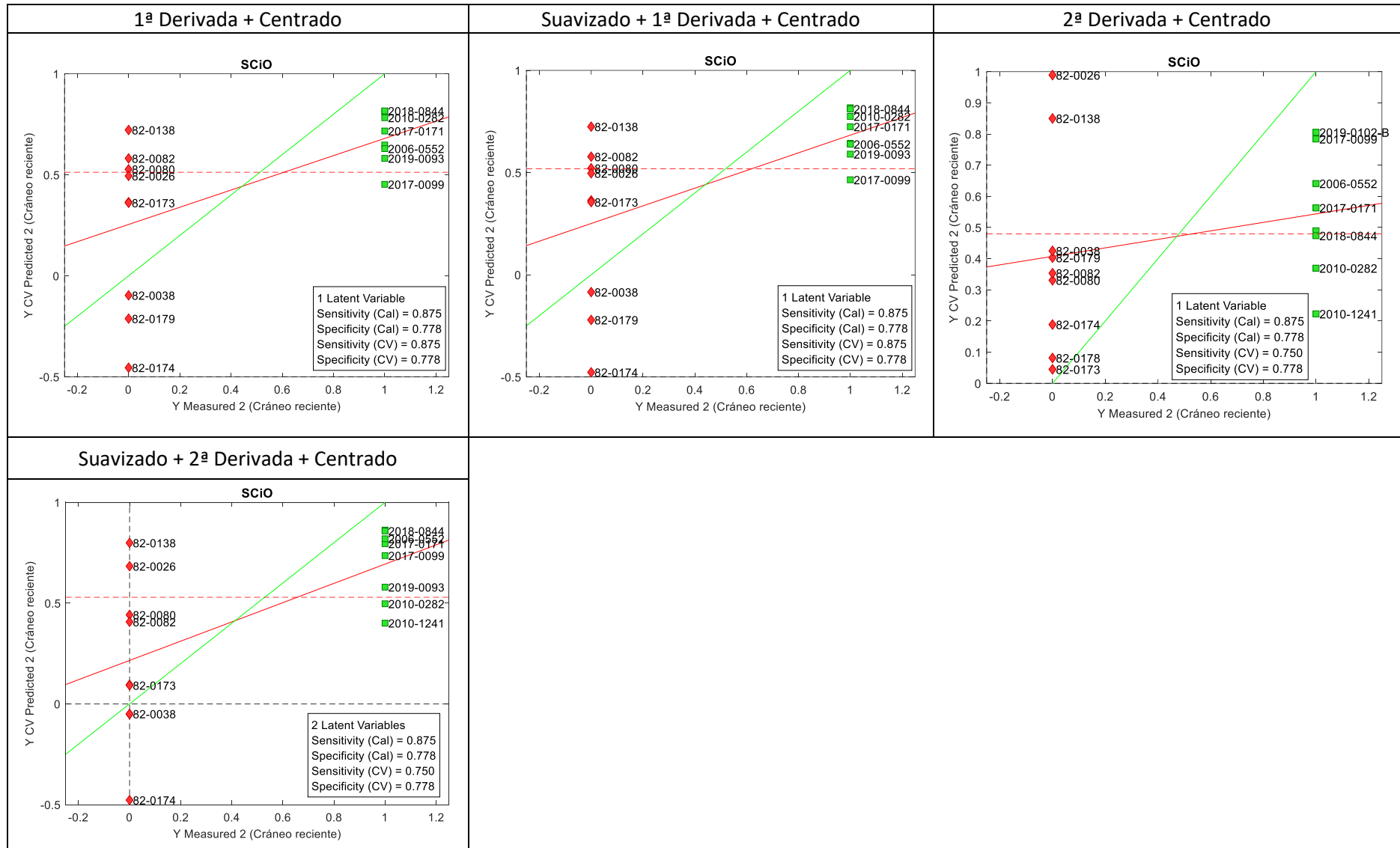
Annexo 2. Pretratamientos con el modelo de clasificación PLS-DA para mandíbulas analizadas con NeoSpectra MDK.





Annexo 3. Pretratamientos con el modelo clasificación PLS-DA para cráneos analizados con SCiO.





Annexo 4. Pretratamientos con el modelo de clasificación PLS-DA para mandíbulas analizadas con SCiO.

