



Universidad de Valladolid

Facultad de Ciencias

TRABAJO FIN DE GRADO

Grado en Física

**Detección Automática de Minerales mediante Espectroscopía Raman.
Acumulación por Multiplicación de Espectros, una Nueva Aproximación.**

Autor:

Elena Pascual Sánchez

Tutor/es:

Fernando Buitrago Alonso, Guillermo López Reyes y Marco Veneranda

Agradecimientos

Gracias a mi tutor, Guillermo López Reyes por acogerme en su equipo y enseñarme lo que significa investigar. Por su constante guía en mi trabajo, sus correcciones y todas sus buenas ideas. Gracias por ser el mejor mentor que hubiera podido tener.

Gracias también a José Antonio, Aurelio y Marco por sus explicaciones y a Mohamed por su ayuda en la parte final de este trabajo.

Le agradezco también su colaboración a Fernando Buitrago, que ha hecho posible la presentación de este estudio.

Y el mayor agradecimiento a mi familia, por su apoyo incondicional e incansable. Nada de esto hubiera sido posible sin vosotros.

Índice

1. INTRODUCCIÓN	5
1.1. OBJETIVOS	5
1.2. CONTEXTUALIZACIÓN DEL TRABAJO	6
1.2.1. Misión ExoMars	6
1.2.2. Instrumento RLS (Raman Laser Spectrometer)	8
1.3. LA ESPECTROSCOPIA RAMAN	10
1.3.1. Fundamento teórico	10
1.3.2. Aspectos prácticos	14
2. TRABAJO EXPERIMENTAL: ADQUISICIÓN DE ESPECTROS RAMAN PARA LA BASE DE DATOS ADAMM Y DE MUESTRAS PROBLEMA PARA LA IDENTIFICACIÓN AUTOMÁTICA.	18
2.1. MATERIALES	18
2.1.1. Espectrómetro de doble pista con láser verde	18
2.1.2. Simulador del instrumento RLS (RLS Sim)	20
2.1.3. Lámparas de calibrado y corrección.	21
2.1.4. Ciclohexano	23
2.2. CALIBRACIÓN DE ESPECTRÓMETROS Y ESPECTROS	24
2.2.1. Calibración del espectrómetro de doble pista	24
2.2.2. Calibración de los espectros adquiridos y creación de SpectrumCalibrator	27
2.2.3. Calibración del RLS Sim	31
2.3. ADQUISICIÓN DE ESPECTROS PARA LA BASE DE DATOS ADAMM	35
2.3.1. Muestras adquiridas	36
2.3.2. Parámetros de adquisición y toma de muestras	38
2.4. ADQUISICIÓN DE ESPECTROS DE MUESTRAS PROBLEMA	41
2.4.1. Muestras adquiridas	41
2.5. CONCLUSIONES DEL TRABAJO EXPERIMENTAL	43
3. TRABAJO COMPUTACIONAL: DETECCIÓN AUTOMÁTICA DE PICOS RAMAN E IDENTIFICACIÓN DE MATERIALES BASADA EN LA MULTIPLICACIÓN DE ESPECTROS.	44
3.1. MATERIALES	45
3.1.1. Espectros del modelo de sustitución de vuelo del instrumento RLS (FS, Flight Spare)	45
3.1.2. Bases de datos de información espectral	46
3.1.3. Software de procesamiento espectral y desarrollo de algoritmia	47
3.1.4. Desarrollo existente del algoritmo de correlación espectral: identificación de minerales en base a picos Raman	47
3.2. DEFINICIÓN DEL MÉTODO DE ACUMULACIÓN POR MULTIPLICACIÓN	51
3.2.1. Estudio preliminar	51
3.2.2. Método optimizado	55
3.3. IMPLEMENTACIÓN DEL ALGORITMO PARA LA DETECCIÓN AUTOMÁTICA DE PICOS	58
3.3.1. Requerimiento de promoción de los picos Raman	60
3.3.2. Método basado en los espectros estadísticos de las muestras de ensayo	61
3.3.3. Sustitución por ceros, media del espectro o redimensión del vector.	63
3.3.4. Método final	64
3.4. IDENTIFICACIÓN AUTOMÁTICA DE MINERALES	73

3.4.1. Adaptación y optimización del algoritmo de clasificación de espectros	73
3.4.2. Definición del algoritmo mejorado para la detección y la clasificación automática	77
3.5. CONCLUSIONES DEL TRABAJO COMPUTACIONAL	83
4. CONCLUSIONES FINALES	84
5. FUTURAS LÍNEAS DE MEJORA	86
6. ANEXO	89
6.1. d_deteccion_automatica_de_picos.m	90
6.2. absoluteMaxima_detection_fn.m	92

Resumen

En el presente trabajo se ha desarrollado un método de identificación automática de materiales basado en su señal Raman. Este método de identificación automática se enmarca en el desarrollo de las herramientas de tratamiento espectral que darán soporte a la misión ExoMars, concretamente al instrumento RLS, desarrollado por el grupo ERICA. El procedimiento implementado tiene como objetivo principal la detección de picos y su clasificación para identificar los minerales o fases asociados a ellos. Dicho procedimiento se sustenta en una metodología novedosa de acumulación de espectros basada en la multiplicación, en lugar de en el tradicional promediado aritmético, acompañado de la definición y parametrización de algoritmos de clasificación de los espectros contra una base de datos espectral.

Para realizar este trabajo, se ha llevado a cabo 1- una investigación teórica de la técnica de espectroscopia Raman; 2- una serie de tareas experimentales orientadas al aprendizaje del manejo de sistemas de adquisición de espectroscopia Raman, que a su vez han permitido generar una base de datos de espectros de referencia y problema que ha sido utilizados para 3- desarrollar las rutinas de identificación automática descritas anteriormente y que son la aportación principal de este trabajo.

Abstract

In this experiment we have developed a new method to automatically classify materials based on their Raman signal. This method of automatic identification has been developed in the framework of the development of the ExoMars mission, specifically, the RLS instrument, whose development is coordinated by the ERICA research group. The implemented procedure will be implemented in the data treatment pipelines for the automatic processing of the RLS data, by detecting the peak information first, and its classification later to identify the minerals or phases associated with those peaks. This routine is sustained on a novel methodology for the accumulation of spectra based on the multiplication of spectra instead of the traditional arithmetic average, together with the definition and parameterization of spectra classification algorithms by comparison with a Raman spectra database.

To carry out this study, we have proceeded as follows: 1-, a theoretical investigation about Raman spectroscopy has been done; 2-, a series of experimental tasks related to the acquisition of Raman spectra allow us to generate a database of reference and problem spectra, which have been used to; 3-, develop the automatic identification routines previously described, which are the main contribution of this work.

1. INTRODUCCIÓN

1.1. OBJETIVOS

El objetivo principal de este trabajo es desarrollar una rutina de análisis espectral que permita clasificar de forma completamente automática un material en base a su señal Raman. Partiendo de una investigación teórica acerca de la espectroscopia Raman, se estudia un método novedoso de acumulación de espectros basado en su multiplicación para mejorar la calidad espectral o relación señal-ruido (SNR). Posteriormente, se desarrolla un algoritmo para identificar automáticamente los picos Raman a partir de los espectros multiplicados, sin detectar falsos positivos. Finalmente, los picos detectados se comparan con la información Raman de diversos espectros presentes en una base de datos, permitiendo la clasificación de la muestra problema.

Adicionalmente, este trabajo también plantea como objetivo la familiarización con el manejo y adquisición de espectros mediante espectroscopia Raman, incluyendo tareas como la calibración de los instrumentos, de los espectros, o la adquisición de espectros de diferentes muestras con condiciones y equipos variados. Además, el trabajo experimental de adquisición de espectros permitirá contar con muestras que se usen posteriormente en el desarrollo de la algoritmia de identificación automática.

1.2. CONTEXTUALIZACIÓN DEL TRABAJO

La identificación automática de materiales será clave al analizar espectros procedentes de Marte, adquiridos por el instrumento RLS de la misión ExoMars de la Agencia Europea del Espacio (ESA). El desarrollo del instrumento RLS (Raman Laser Spectrometer) ha sido coordinado por el grupo de investigación ERICA (Espectroscopia Raman e Infrarroja aplicada a Cosmogeología y Astrobiología) de la Universidad de Valladolid y se encuentra integrado en el rover Rosalind Franklin de la ESA. Sus características principales, así como los objetivos de la misión se detallan a continuación.

1.2.1. Misión ExoMars

La misión ExoMars (Exobiology on Mars) de la ESA y la agencia rusa Roscosmos es una misión de carácter astrobiológico cuyo objetivo es hallar indicios de vida, tanto presente como pasada, en Marte. A su vez, pretende caracterizar el entorno atmosférico y geológico del planeta rojo, con especial énfasis en la presencia de agua.

La misión cuenta con tres partes diferenciadas: un satélite para la detección de gases traza en la atmósfera marciana (Trace Gas Orbiter, TGO), un rover que se desplazará por la superficie a partir de su llegada en 2023 y una plataforma de aterrizaje o lander, que ayudará al rover a amartizar de forma segura y que está equipado con tecnología para llevar a cabo experimentos en suelo marciano.

La gran novedad que presenta Rosalind Franklin con respecto a sus compañeros en Marte es que será el primer rover que taladre hasta dos metros de profundidad en el subsuelo marciano, de modo que recogerá muestras que han estado protegidas de la radiación cósmica (ionizante y no ionizante) y la erosión superficial, aumentando las posibilidades de detección de materia orgánica por su mayor grado de protección frente a agentes externos. Las muestras recogidas a esa profundidad serán pulverizadas y analizadas en el laboratorio analítico del rover, o Analytical Laboratory Drawer (ALD), por los instrumentos MICROMEGA, RLS y MOMA:

- MICROMEGA: Espectrómetro infrarrojo y visible para el estudio de minerales.
- RLS: (Raman Laser Spectrometer) Espectrómetro Raman para la identificación de compuestos minerales.
- MOMA (Mars Organic Molecule Analyser): Cromatógrafo de gases/desorbición láser y espectrómetro de masas. Operará dentro de PASTUR y también analizará gases de la atmósfera Marciana.

Para dar soporte a la misión y adquirir información complementaria, el rover también cuenta con los siguientes instrumentos⁽¹⁾ que, en combinación con los instrumentos del laboratorio analítico, conforman la carga de pago PASTEUR del rover Rosalind Franklin:

- PANCAM (PANoramic CAMera): Cámara panorámica ubicada en el mástil del rover que ofrecerá información para la navegación y la selección de lugares de interés para efectuar campañas de ciencia.
- ISEM (Infrared Spectrometer for ExoMars): Espectrómetro infrarrojo para la detección de compuestos.
- CLUPI (CLOse UP Imager) Cámara para la obtención de imágenes de alta resolución en color.
- WISDOM (Water Ice and Subsurface Deposit Observation On Mars): Radar para la caracterización de la estratigrafía bajo el rover. Ayudará a la selección de lugares para la perforación aportando información acerca de su contenido acuoso.

- ADRON-RM (Autonomous Detector of Radiation Onboard Rover at Mars) Espectrómetro de neutrones para la búsqueda de hielo y minerales hidratados.
- MA-MISS (Mars Multispectral Imager for Subsurface Studies) Espectrómetro visible e infrarrojo situado en el mástil del rover.

Con el objetivo de hallar signos de vida tanto presente como pasada en Marte, Oxia Planum(2) ha sido seleccionado, de entre cinco candidatos, como lugar de aterrizaje para la misión. La decisión fue tomada en base a los requerimientos científicos y la necesidad de llevar a cabo un amartizaje seguro. La región de Oxia Planum es una llanura de relieve suave constituida principalmente por arcillas. Su terreno data de la era Noéica (ver Figura 1.) de la historia marciana y se tiene constancia de que ha albergado ecosistemas acuosos en al menos dos ocasiones: una primera, que dio lugar a la sedimentación de materiales arcillosos en capas y una segunda, en la cual la zona se comportó como un delta fluvial que diseminó la estructura de capas previamente formada.

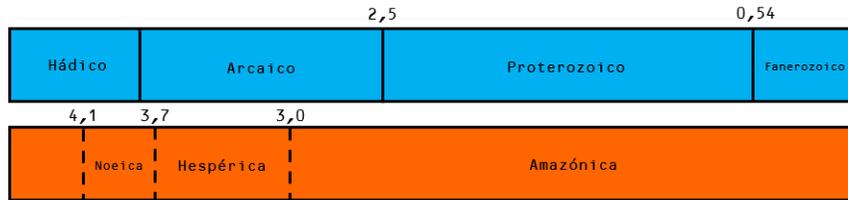


Figura 1: Eras de la historia terrestre y marciana(3). Edades en miles de Ma.

Debido a su antigüedad, la evidencia de un pasado acuoso y la abundancia de materiales a lo largo del terreno, esta zona, ubicada entre 16°-19° N y -23°- -28°E, resulta prometedora para la búsqueda de trazas biológicas.

1.2.2. Instrumento RLS (Raman Laser Spectrometer)

El instrumento RLS(4) fue concebido como uno de los primeros espectrómetros Raman destinados a la exploración espacial. Pesa tan solo 2,483kg y cuenta con tres partes diferenciadas:

- **SPU (Spectrometer Unit)**

La unidad del espectrómetro dispone de un sistema óptico para el colimado y la recolección de la luz, una red de difracción holográfica y una CCD (charge-coupled device) refrigerada que actúa como detector. Está basado en una configuración óptica de transmisión (véase Figura 2.) que permite maximizar la eficiencia en términos de luz recogida.



Figura 2: Espectrómetro (SPU) del instrumento RLS.

- **iOH (Optical Head Unit)**

El cabezal óptico del instrumento (Figura 3.) tiene la función de enfocar el haz láser sobre la muestra, así como recoger la luz dispersada que después se guiará hasta el espectrómetro. La anchura del haz incidente es de $50\mu m$. El sistema de enfoque permite salvar las irregularidades de los granos minerales de la muestra. En la región de recolección de luz, el cabezal cuenta con un filtro para eliminar la dispersión Rayleigh.

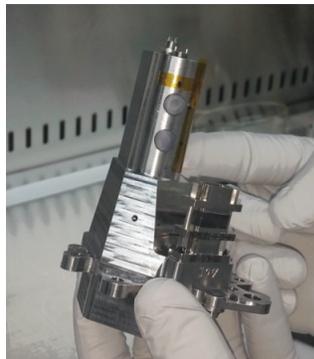


Figura 3: Cabezal óptico (iOH) del instrumento RLS.

■ **ICEU (Instrument Control and Electronics Unit)**

La unidad de control electrónico incorpora la fuente de excitación láser ($\lambda = 532nm$), la unidad de suministro de potencia, un procesador para el control del equipo, el guardado de datos y la comunicación con el rover y la unidad de comunicación con la CCD que transmite los datos al procesador.

Láser del instrumento RLS

El diodo láser del instrumento RLS es un láser de estado sólido preparado para operar entre 15°C y 45°C, con funcionamiento óptimo a 25°C. Su emisión es muy estable (variación en la posición de la longitud de onda: $\pm 50pm$). Para obtener la emisión láser deseada se ilumina un cristal de granate de itrio y aluminio dopado con neodimio al 1%, comúnmente conocido como cristal de Nd:YAG, con un láser de 808nm. El cristal emite a 1064nm, frecuencia que se divide a la mitad con un cristal generador de segundos armónicos (SHG) obteniéndose finalmente una emisión de 532nm.

La longitud de onda de 532nm es muy adecuada para la misión puesto que posee un gran rendimiento y estimula las señales de resonancia Raman de las biomoléculas.

El láser está preparado para suministrar una potencia óptica de 20mW.

Otros elementos importantes del instrumento son su muestra de calibración (CT, Calibration Target), así como las conexiones ópticas y eléctricas. Todo el equipo está preparado para soportar las bajas temperaturas e inclemencias del ambiente marciano, así como los requerimientos mecánicos y térmicos del lanzamiento, crucero y amartizaje.

1.3. LA ESPECTROSCOPIA RAMAN

La espectroscopia Raman (5) se basa en la dispersión inelástica de la radiación electromagnética. En su interacción, los fotones y las vibraciones moleculares intercambian energía. De entre todos los fotones que sufren la dispersión al interactuar con una molécula, solo 1 ó 2 por cada 10^{10} posee una frecuencia distinta a la radiación incidente. Esta radiación desplazada en longitud de onda es característica de los enlaces moleculares con los que se produce la interacción, de forma que los desplazamientos que se producen en los fotones reemitidos son característicos de la molécula. Por tanto, mediante la identificación de las longitudes de onda emitidas por los fotones Raman, es posible realizar la identificación molecular del material expuesto a la radiación. Esta característica hace que el análisis mediante espectroscopia Raman sea muy potente en cuanto a la capacidad de identificación de materiales, ya que permite distinguir incluso casos de polimorfismo (materiales con la misma fórmula química pero diferente estructura molecular).

Desde su descubrimiento en 1928, la espectroscopia Raman ha tenido un desarrollo muy positivo, acompañado en los últimos años de una gran aceptación gracias a los avances técnicos, pero sobre todo a causa de sus numerosas ventajas analíticas entre las que destacan:

- Es una técnica no destructiva, que no requiere preparación previa de la muestra.
- Puede usarse de forma remota sin contacto directo con el material a analizar, así como in situ gracias a la miniaturización de los equipos y la existencia de espectrómetros Raman portátiles.
- Ofrece una precisión elevada en la identificación de muestras ya que es capaz de identificar la composición, el grado de orden, la estructura del material, la presencia de impurezas, la semicuantificación de abundancias minerales en muestras, etc. Su alta especificidad justifica su favorable desarrollo.

La información que se muestra en la presente sección procede de los libros P. Vandenabeele, 2013, *Practicas Raman Spectroscopy - An Introduction*, Universidad de Ghent, Bélgica, John Wiley & Sons Ltd. (5) y F. Rull, comp., 1993, *Espectroscopia Infrarroja y Raman: Fundamentos y Aplicaciones en Mineralogía*, Valladolid, España, Universidad de Valladolid (6).

1.3.1. Fundamento teórico

Aproximación clásica al efecto Raman Cuando una molécula se sitúa en un campo eléctrico \vec{E} , se induce en ella un momento dipolar \vec{p} . La relación existente entre el campo eléctrico y el momento dipolar inducido puede expresarse como desarrollo en serie de potencias:

$$\vec{p} = \alpha\vec{E} + \frac{1}{2}\beta\vec{E}^2 + \frac{1}{6}\gamma\vec{E}^3 + \dots \quad (1)$$

donde α es el tensor polarizabilidad, β el tensor hiperpolarizabilidad y γ el tensor segunda hiperpolarizabilidad. Los términos a partir de segundo orden acostumbran a ser ignorados por ser 10 órdenes de magnitud menores que su término predecesor en la serie.

En consecuencia, la expresión para el momento dipolar inducido se simplifica a

$$\vec{p} = \alpha\vec{E} \quad (2)$$

Al estudiar la dispersión Raman, el campo eléctrico que induce el momento dipolar procede de la radiación electromagnética, que podemos describir como:

$$\vec{E} = \vec{E}_0 \cos(2\pi\nu_0 t) \quad (3)$$

Coefficiente	Valor típico
α	$10^{-40} CV^{-1} m^2$
β	$10^{-50} CV^{-2} m^3$
γ	$10^{-60} CV^{-3} m^4$

Tabla 1: Valores típicos para los coeficientes polarizabilidad, hiperpolarizabilidad y segunda hiperpolarizabilidad

siendo ν_0 la frecuencia de dicha radiación.

El tensor polarizabilidad, a su vez, da cuenta de la relación existente entre las vibraciones del campo eléctrico y los fonones o vibraciones de la red. Depende de la forma y el tamaño del enlace químico y será sensible a las modificaciones que sufra este durante las vibraciones de la molécula. Para recoger estos efectos se expande el tensor polarizabilidad en serie de Taylor en función de las coordenadas normales, Q , de la molécula:

$$\alpha = \alpha_0 + \sum_k \left(\frac{\partial \alpha}{\partial Q_k} \right)_0 Q_k + \frac{1}{2} \sum_{k,l} \left(\frac{\partial^2 \alpha}{\partial Q_k \partial Q_l} \right)_0 Q_k Q_l + \dots \quad (4)$$

Q_k y Q_l son las coordenadas normales correspondientes a la k -ésima y l -ésima vibración normal, a las cuales se corresponden unas frecuencias ν_k y ν_l , respectivamente. Para continuar con el desarrollo, consideraremos una sola vibración independiente e ignoraremos los términos cruzados, de manera que el tensor polarizabilidad a ella asociado tendrá la forma

$$\alpha_\nu = \alpha_0 + \alpha'_\nu Q_\nu \quad (5)$$

Donde α'_ν es la derivada del tensor polarizabilidad bajo condiciones de equilibrio. En primera aproximación, las coordenadas normales vibrarán de forma armónica, por lo que pueden ser descritas:

$$Q_\nu = Q_{\nu_0} \cos(2\pi\nu_\nu \cdot t + \phi_\nu) \quad (6)$$

Reescribiendo ahora la expresión del momento dipolar simplificada:

$$\vec{p} = \alpha \vec{E} = \alpha_0 \vec{E}_0 \cos(2\pi\nu_0 t) + \alpha'_0 Q_{\nu_0} \vec{E}_0 \cos(2\pi\nu_0) \cos(2\pi\nu_\nu t + \phi_\nu) \quad (7)$$

y empleando la relación trigonométrica

$$\cos(\theta_1) \cos(\theta_2) = \frac{1}{2} [\cos(\theta_1 + \theta_2) + \cos(\theta_1 - \theta_2)] \quad (8)$$

llegamos a la expresión final

$$\vec{p} = \alpha \vec{E} = \alpha_0 \vec{E}_0 \cos(2\pi\nu_0 t) + \frac{1}{2} \alpha'_0 Q_{\nu_0} \vec{E}_0 \cos(2\pi(\nu_0 + \nu_\nu)t + \phi_\nu) + \frac{1}{2} \alpha'_0 Q_{\nu_0} \vec{E}_0 \cos(2\pi(\nu_0 - \nu_\nu)t + \phi_\nu) \quad (9)$$

Donde se relacionan la frecuencia de vibración de la molécula ν_ν , con la frecuencia de la radiación electromagnética ν_0 . El momento dipolar inducido puede separarse en tres componentes:

(10)

$=(\nu_0) + \bar{p}(\nu_0 + \nu_\nu) + \bar{p}(\nu_0 - \nu_\nu)$ La primera de ellas se corresponde con la dispersión elástica o dispersión Rayleigh, mucho más intensa que la dispersión Raman. El segundo y tercer término se corresponden con la dispersión Raman de mayor energía o Anti-Stokes y la de menor energía o Stokes, respectivamente (véase Figura 4).

Al llevar a cabo un experimento de dispersión Raman, un haz monocromático, típicamente un láser, incide sobre la muestra. La radiación dispersada se mide en función de su longitud de onda. Sin embargo, para guardar independencia con la longitud de onda de excitación suele representarse la intensidad de radiación dispersada, en cuentas, en función del desplazamiento Raman, medido como diferencia de frecuencias entre la luz incidente y la dispersada, en cm^{-1} .

$$D_R = \frac{\nu_d - \nu_0}{c} \quad (11)$$

donde D_R denota el desplazamiento Raman, ν_d , la frecuencia de la radiación dispersada y ν_0 , la frecuencia de la radiación incidente.

Reglas de selección Considerando la aproximación armónica para los modos normales, puede verse que la energía de las señales Raman estará determinada por la diferencia existente entre el estado fundamental y el primer estado excitado.

Gracias a esta aproximación, la frecuencia de la vibración puede describirse como

$$\nu_\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (12)$$

Donde k es la constante de fuerza del enlace y μ , la masa reducida. La posición de la señal Raman dependerá de las características concretas del enlace, así como de las masas de los átomos que una. Estos parámetros son bastante estables para cada tipo de enlace o grupo funcional y por tanto la señal Raman proporciona una huella dactilar muy característica para cada uno de ellos que permite una identificación precisa del material bajo estudio. Sin embargo, esta ecuación no indica si un modo es activo Raman o no, por lo que debe ser combinada con la regla de selección previamente explicada. En la figura 4. pueden verse los esquemas energéticos de los distintos tipos de dispersión. Según el modelo aproximado del oscilador armónico, los distintos niveles están separados una cantidad $\frac{1}{2}\hbar\nu_\nu$.

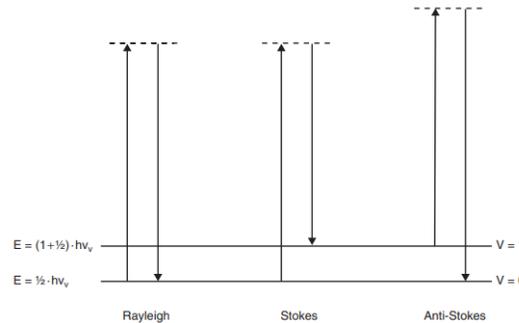


Figura 4: Esquema energético de los distintos tipos de dispersión.

Intensidad de la radiación Raman Para estimar la intensidad Raman de un espectro, deben considerarse dos grupos de efectos:

$$I = \epsilon \cdot \tau \quad (13)$$

Los efectos relacionados con la muestra ϵ , y los relacionados con el instrumento empleado τ . Los efectos relacionados con el instrumento incluyen la intensidad de la fuente de radiación, la eficiencia del detector, la geometría de la medida, etc. Si consideramos la intensidad desde un punto de vista fundamental, ignorando las contribuciones del instrumento, podremos ver que guarda una fuerte dependencia con la orientación en la que el efecto Raman es medido.

En general, la intensidad de la luz dispersada es igual a la variación sufrida en el flujo luminoso bajo un cierto ángulo sólido

$$I = \frac{d\Phi}{d\Omega} = \frac{\pi^2 c \bar{\nu}^4 p_0^2 \text{sen}^2(\theta)}{2\epsilon_0} \quad (14)$$

Donde $\bar{\nu}$ es el número de onda absoluto de la radiación electromagnética, p_0 , la amplitud del momento dipolar inducido, c , la velocidad de la luz en el vacío y ϵ_0 , la permitividad eléctrica del vacío. θ es el ángulo existente entre la dirección del momento dipolar y la dirección de medida. La expresión propuesta es una simplificación pues la medida no se efectúa en una dirección concreta sino bajo un cierto ángulo sólido, pero es útil para mostrar la fuerte dependencia que tiene la intensidad con la dirección de medida (dependencia $\text{sen}^2(\theta)$) y con el número de onda absoluto de la radiación electromagnética. Si la intensidad es proporcional a $\bar{\nu}^4$, será también inversamente proporcional a λ^4 , con lo que a menor longitud de onda empleada como fuente de excitación, mayor intensidad Raman se obtendrá en los espectros.

Anchura de la señal Raman La anchura de los picos Raman posee distintas contribuciones, la principal de ellas se debe a los vecinos que rodean a la molécula y que alteran las constantes de fuerza de los enlaces que la caracterizan. Esto explica porque los materiales amorfos poseen señales mucho más anchas que aquellos que son cristalinos. Otros cambios destacados se relacionan con la temperatura, el cambio de cristalinidad de la muestra o la variación en las masas del enlace.

1.3.2. Aspectos prácticos

Fluorescencia en espectros Raman Cuando una molécula absorbe la radiación incidente y se promociona a un estado electrónico excitado, tiene lugar el fenómeno de fluorescencia. Mediante un decaimiento sin emisión de radiación la molécula regresa a un estado de menor energía desde el que puede alcanzar el estado fundamental, esta vez, con emisión de radiación (véase Figura 5.).

Esta emisión de radiación, si bien tiene tiempos de recombinación órdenes de magnitud mayores que aquellos de la emisión Raman, interfiere con la señal Raman, pudiendo enmascararla en muchas ocasiones cuando se utilizan equipos de medida con láseres de emisión continua.

Para disminuir el efecto de la fluorescencia en el espectro puede llevarse a cabo un proceso de photobleaching o quenching, que consiste en irradiar la muestra durante un periodo prolongado de tiempo para desactivar o saturar los fluoróforos, respectivamente.

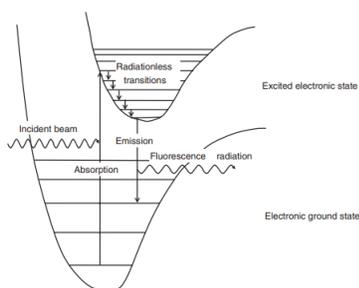


Figura 5: Esquema energético del proceso de fluorescencia.

Ruido en espectroscopia Raman Se define por ruido a aquellas señales presentes en el espectro que no proporcionan información relevante para el experimento, y que típicamente afectarán de forma aleatoria o pseudo-aleatoria a las señales medidas por el experimento. Aunque hay muchas posibles, las fuentes de ruido son comúnmente tres:

- **Ruido asociado a la señal, σ_s**

El ruido asociado a la señal o ruido shot se pone de manifiesto cuando las fluctuaciones estadísticas son apreciables. Por tanto, este ruido será más notable de forma inversamente proporcional al tiempo de observación, es decir, el tiempo de adquisición. La contribución marginal del ruido shot será mayor cuanto más elevada sea la intensidad del espectro, ya que es proporcional a la raíz cuadrada de la misma. Sin embargo, la contribución neta de la SNR será menor cuanto mayor sea la intensidad.

- **Ruido procedente de la instrumentación empleada, σ_i**

A causa de la temperatura, en el detector del instrumento se generan de forma espontánea pares electrón-hueco sin que incida luz sobre él. Para evitar esta generación espontánea que aumenta el ruido del espectro, conviene refrigerar el detector. La refrigeración mejora muy notablemente la relación señal-ruido, sin embargo, una cierta contribución se mantiene. Para lograr eliminarla de una forma más exhaustiva se toma un espectro con igual tiempo de integración que el espectro problema, pero con el láser apagado y se sustrae al espectro original. De este modo se elimina la contribución térmica característica del detector. La sustracción del espectro en oscuridad también

elimina la contribución de la luz ambiental. Este proceso suele denominarse substracción del dark o del background.

■ **Ruido causado por el procesado de los espectros, σ_p**

Los componentes electrónicos que se encargan de transformar la señal analógica en digital también pueden inducir ruido en el espectro.

La contribución final de estos ruidos al espectro es la suma cuadrática de todas las componentes de ruido, que quedaría por tanto definido según la siguiente ecuación:

$$\sigma = \sqrt{\sigma_s^2 + \sigma_i^2 + \sigma_p^2} \quad (15)$$

SNR La relación señal-ruido o Signal to Noise Ratio (SNR) es, ortodoxamente hablando, la variabilidad de la señal a lo largo de diferentes espectros tomados con las mismas condiciones de adquisición. Sin embargo, dicha definición no proporciona una idea de la calidad espectral de un espectro (cómo de bien se pueden detectar los picos de emisión Raman con respecto al fondo del espectro). Por ello, está ampliamente extendida una definición alternativa de SNR que en se ha definido para este trabajo como el valor de intensidad de un pico Raman (I_R) medido desde su base dividido por la desviación estándar del espectro en un rango de 200cm^{-1} (std_{200}) que no presente información Raman.

$$SNR = \frac{I_R}{std_{200}} \quad (16)$$

De esta forma, el valor de SNR proporcionará un valor relacionado con la capacidad de distinguir un pico de emisión de las variaciones del fondo del espectro.

Adquisición de espectros Raman Debido a la variabilidad en la emisión Raman de las diferentes muestras que se puedan analizar, el procedimiento típico que se lleva a cabo para adquirir un espectro Raman requiere optimizar los parámetros de adquisición para adaptarse a la muestra bajo análisis. Para ello, en espectrómetros Raman con CCD refrigerada y alta luminosidad como el instrumento RLS, el procedimiento óptimo consiste en tomar el espectro con el mayor tiempo de acumulación posible que permita el espectrómetro sin saturar el detector. Esto se hace así ya que la relación entre el tiempo de acumulación y la intensidad del espectro es lineal, mientras que el crecimiento del ruido no lo es. Como se ha mostrado en la literatura (7), cuanto más tiempo de integración se emplee para adquirir el espectro, más se promocionará la señal Raman frente al ruido espectral. La relación existente es lineal (véase Figura 6.). El límite de saturación está marcado por las cuentas que es capaz de registrar el espectrómetro.

Para mejorar aún más la relación señal-ruido del espectro se realiza el procedimiento conocido como acumulación de espectros, de forma que se toman varios espectros del mismo punto de la muestra bajo iguales condiciones de tiempo de integración. El procedimiento habitual de acumulación de espectros estándar que se lleva a cabo pasa por consiste en sumar estos espectros (o promediarlos aritméticamente si se divide entre el número de acumulaciones) y tratar de forma que el espectro suma como uno solo acumulado obtenga una mejor SNR al cancelar matemáticamente las contribuciones aleatorias del ruido de los espectros individuales.

En los casos en los que la muestra presenta una fluorescencia elevada, es conveniente esperar un tiempo iluminando el punto de estudio con el haz láser para que tenga lugar el fenómeno de photo-

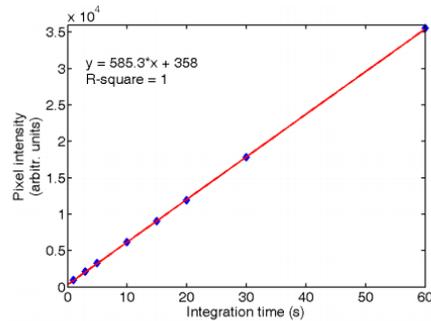


Figure 3-40. Intensity at one CCD pixel vs. Integration time

Figura 6: Intensidad de cada píxel del espectro en función del tiempo de integración. La relación existente es lineal. (7).

bleaching o quenching de la fluorescencia.

La adquisición de muestras ha de llevarse a cabo en un entorno oscuro para que la luz ambiental, especialmente la que procede de lámparas artificiales, no contribuya al espectro confundiendo la información Raman. No obstante, frente a la imposibilidad de tener un espacio en absoluta oscuridad, es habitual realizar la resta de los darks o espectros de oscuridad, adquiridos en las mismas condiciones de medida pero sin activar la fuente de excitación. Esta corrección permitirá no solo eliminar la contribución de la luz ambiental parásita, sino que también permitirá la corrección o eliminación de aquellos efectos de ruido o bias que no presenten un comportamiento aleatorio.

Corrección en intensidad de espectros Según *J.A. Sanz* (8), los equipos de espectroscopia Raman cuentan con un cabezal para el enfoque y recolección de la luz. Este cabezal se conecta mediante fibra óptica al espectrómetro donde el haz es separado ópticamente mediante la red de difracción, de forma que en sus distintas componentes frecuenciales de la luz emitida son para ser recogidas espacialmente distribuidas en una CCD. Sin embargo, todos Estos elementos ópticos introducen un patrón característico en los espectros denominado función aparato que afecta a la intensidad de la señal de forma diferente en función de la longitud de onda de la emisión (por ejemplo, el CCD puede ser más eficiente a la hora de captar los fotones en una determinada longitud de onda frente a otras). La función aparato se convoluciona con la emisión Raman “pura”, afectando a la forma incluye como contribución a de la línea de base del espectro, dando lugar a una curva no perfecta que en la práctica puede presentar formas de lo más variopintas.

La división entre un espectro amorfo de calibración permite llevar a cabo la corrección de intensidad absoluta del instrumento. Diversos patrones comerciales permiten llevar a cabo esta corrección. Algunos de ellos son:

- Lámparas de tungsteno calibradas que presentan una emisión muy bien caracterizada.
- Estándares luminosos de calibración como los proporcionados por el NIST (National Institute of Standards and Technology).

Para crear una solución casera y accesible, el grupo de investigación ERICA se propuso, basándose en su experiencia trabajando en espectroscopia Raman, hallar un material que permitiera corregir

las imperfecciones inducidas en la línea de base con recursos de fácil acceso. Para ello se investigó la composición del estándar de calibración NIST: cristal de borosilicato dopado con cromo y se buscaron materiales de características similares, para finalmente seleccionar el borato de zinc como candidato a patrón de calibración casero. El borato de zinc $Zn[B_3O_4(OH)_3]$ es un compuesto usado en industria como aditivo para polímeros, cerámicas y otras aplicaciones. A partir de 560k comienza a deshidratarse y a 690K finaliza su deshidratación dando lugar a una fase amorfa ($2ZnO \cdot 3B_2O_3$) perfecta para proporcionar una señal ancha sin picos de emisión Raman que pueda corregir la función aparato.

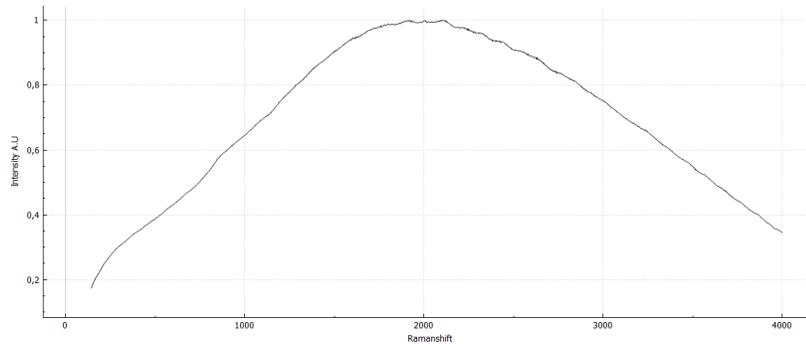


Figura 7: Espectro de borato de zinc adquirido con el modelo de sustitución de vuelo RLS-FS.

2. TRABAJO EXPERIMENTAL: ADQUISICIÓN DE ESPECTROS RAMAN PARA LA BASE DE DATOS ADAMM Y DE MUESTRAS PROBLEMA PARA LA IDENTIFICACIÓN AUTOMÁTICA.

La primera parte del trabajo consiste en la adquisición de espectros Raman de muestras seleccionadas para la base de datos ADAMM (14) que se utilizará para la identificación espectral, así como de muestras problema empleadas para verificar y testear su identificación automática. Este proceso requiere los siguientes materiales.

2.1. MATERIALES

Los materiales requeridos para la adquisición de los espectros son:

- Espectrómetro de doble pista con láser verde.
- Simulador del instrumento RLS.
- Lámparas de calibrado y corrección
- Ciclohexano

2.1.1. Espectrómetro de doble pista con láser verde

El espectrómetro de doble pista empleado es un espectrómetro de la familia HoloSpec VPHTM System(9) de la empresa Kaiser Optical Systems, INC. Este espectrómetro es un espectrómetro compacto de gran resolución.

La configuración óptica del sistema es tal que la salida del espectrómetro está conectada con una cámara refrigerada cuya CCD cuenta con 2048 píxeles (puede verse en la Figura 8.). El proceso de filtrado del pico de dispersión Rayleigh se produce en el cabezal Raman, pues el espectrómetro no cuenta con una pre-etapa de filtrado a la entrada, emulando así la configuración del instrumento RLS.

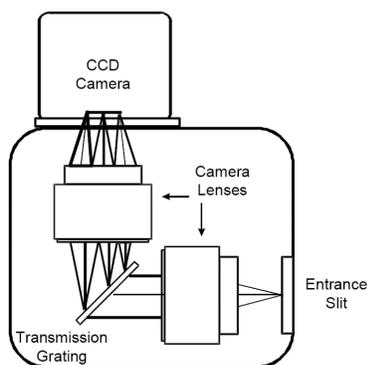


Figura 8: Configuración óptica del espectrómetro de doble pista.

La configuración óptica del sistema es en retrodispersión, es decir, la radiación que ha sufrido la dispersión Raman es recogida en el mismo cabezal por el cual se emite el haz láser.

El espectrómetro ha sido empleado con dos láseres diferentes:

- Láser Oxixius. Láser de longitud de onda $\lambda = 532nm$ que proporciona $53mW$ a su salida.
- RAD2. Láser de $\lambda = 532nm$ que proporciona $60mW$ a su salida.

Así como con dos cabezales Raman distintos:

- BWTEK-2011 (“RAD-2”) Cabezal Raman de menor luminosidad.
- BWTEK-2008 (“i-Raman”) Cabezal Raman de mayor luminosidad.

La combinación empleada con la cual se obtenía mayor luminosidad era acoplando el láser RAD-2 al cabezal BWTECK-2008, llegando a obtener $48mW$ de potencia sobre la muestra. Debido a requerimientos experimentales, fue necesario, durante un tiempo, volver a emplear el cabezal BWTECK-2011, trabajando con una potencia de $34,5mW$. Ambas configuraciones pueden observarse en la Figura 9.

El espectrómetro es un espectrómetro de doble pista. Los espectrómetros de doble pista se caracterizan porque su red de difracción holográfica está doblemente grabada. Esto hace que la luz se difracte en dos regiones diferentes dependiendo de su longitud de onda. Esto se traduce en que el equipo tiene el doble de resolución que si funcionara con una red sencilla. En el detector del equipo, su CCD, la luz dispersada por cada pista de la red se recoge en regiones diferentes.

El detector del espectrómetro es una cámara comercializada por *ANDORTM TECHNOLOGY*, modelo newton. El equipo cuenta con un sistema de refrigeración que permite disminuir la temperatura de la CCD hasta $-90^{\circ}C$. Al refrigerar el detector se disminuye la contribución del ruido térmico y pueden tomarse espectros con mayor tiempo de integración, con la consiguiente mejora de la relación señal-ruido.

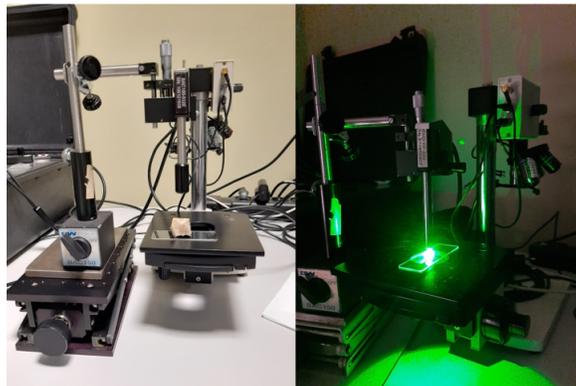


Figura 9: Distintos cabezales Raman acoplados por fibra óptica al espectrómetro de doble pista.

2.1.2. Simulador del instrumento RLS (RLS Sim)

El simulador del instrumento RLS o RLS Sim, ubicado en la sede del grupo de investigación ERICA, en el Parque Tecnológico de Boecillo (véase Figura 10.), emula el funcionamiento del instrumento a bordo del rover Rosalind Franklin. El simulador cuenta con elementos de hardware: espectrómetro, láser, fibras ópticas, luces de posicionamiento, platina de muestras y sistema de distribución que reproducen lo más fielmente posible las características del instrumento de vuelo. A su vez, el software de control del instrumento, programado en LabView, permite llevar a cabo rutinas de adquisición automática de espectros como las que se llevarán a cabo en Marte. De este modo es posible analizar de forma autónoma múltiples puntos de diversas muestras. Durante dichas rutinas diversos parámetros pueden ser modificados:

- Potencia del láser: Seleccionada al 70 % o al 90 % dependiendo del láser utilizado, para obtener los 20mW aproximados con los que opera el láser del modelo de vuelo.
- Número de puntos a analizar y distribución de los mismos: Para las muestras estudiadas se han seleccionado 6 puntos por muestra.
- Número de acumulaciones y N_{av} : El número de acumulaciones da cuenta de la cantidad de espectros adquiridos sobre un mismo punto de la muestra. A su vez, N_{av} indica la cantidad de espectros empleados para el promedio (aritmético). N_{av} es por defecto 0, de manera que todos los espectros adquiridos se integran en uno solo. Sin embargo, al seleccionar $N_{av} = 1$, se obtienen cada uno de los espectros por separado, permitiendo después llevar a cabo con ellos el tratamiento deseado, que en el caso de este trabajo consiste en la multiplicación de los mismos, como se explicará en la Parte 2. Este parámetro está integrado también a bordo del rover, quien lo usará para mandar la mayor cantidad de datos posibles en función del ancho de banda que le permitan los satélites que establecen la comunicación con la Tierra.

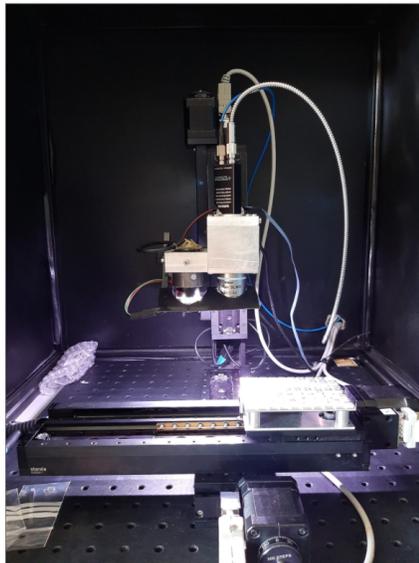


Figura 10: Simulador RLS ubicado en su cámara de oscuridad. La platina de muestras se encuentra en el punto de manipulación.

Durante el desarrollo de la rutina automática llevada a cabo por el espectrómetro, se efectúa un correcto posicionamiento sobre la muestra, el enfoque automático de un punto la misma, la reducción de la fluorescencia, la adquisición del espectro, la adquisición de un dark con igual tiempo de integración, la eliminación del dark y la muestra de resultados.

El simulador también permite su operación manual.

2.1.3. Lámparas de calibrado y corrección.

Para que la información adquirida por los espectrómetros en los píxeles de su CCD pueda traducirse a desplazamientos Raman, es necesario calibrar estos equipos. Para ello, pueden emplearse lámparas de calibración o materiales con una emisión Raman muy bien conocida y estable como los estándares establecidos por la ASTM. Mediante lámparas de corrección también es posible eliminar la contribución de la función aparato a la intensidad de los espectros.

Las lámparas empleadas para la calibración de los espectrómetros son:

- Lámpara de gas de Mercurio y Argón de la compañía Ocean Optics. La lámpara cubre un rango espectral de 253 a 1700nm. La salida de la lámpara ha de conectarse a una fibra óptica de 50µm de diámetro o menos. El espectro de emisión de esta lámpara puede observarse en la Figura 11.

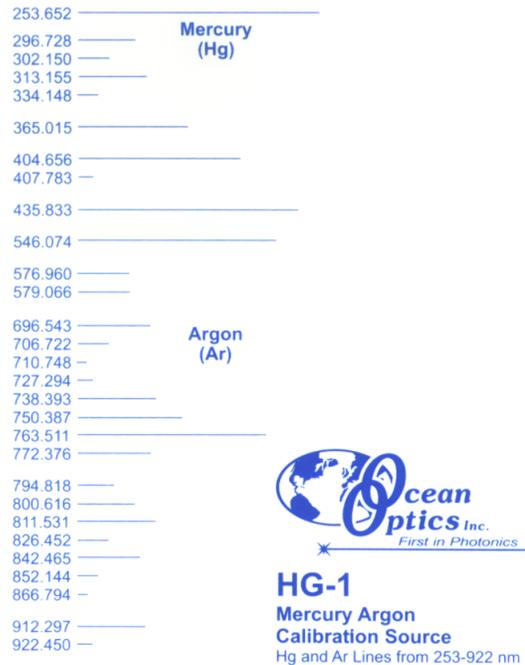


Figura 11: Espectro de emisión de la lámpara de Mercurio-Argón. La longitud de onda de las señales está expresada en nm.

- Lámpara de Neón perteneciente al set Raman Calibration Accessory(10) de la compañía Kaiser Optical Systems, Inc. Las líneas de emisión atómica del Neón cubren densamente el rango espectral bajo estudio, ya que poseen gran cantidad de picos a lo largo de este como puede verse en la Figura 12. Concretamente, en el caso en el que se trabajará su emisión permite calibrar la zona de elevados desplazamientos Raman de los espectrómetros.

533.0778	607.4338	638.2992	743.8900	783.9055	920.1760
540.0562	609.6163	650.6528	748.8871	794.3181	937.3310
556.2766	614.3063	659.8953	753.5774	808.2458	948.6680
574.8298	626.6495	671.7043	754.4044	830.0326	966.5420
588.1895	638.2992	692.9467	794.3181	841.8427	1029.5420
594.4834	650.6528	703.2413	808.2458	859.1259	1056.2410
597.5534	667.8276	717.3938	830.0326	891.9501	1079.8070
607.4338	671.7043	724.5167	837.7606	914.8670	
609.6163	692.9467		841.8427	920.1760	

Figura 12: Espectro de emisión de la lámpara de Neón. La longitud de onda de las diferentes señales está expresada en nm.

A su vez, para llevar a cabo la corrección en intensidad una vez que los espectros han sido adquiridos, se emplea una lámpara de emisión ancha muy bien caracterizada: Modelo HCA (HoloLab Calibration Accessory) de la marca Kaiser Optical Systems, Inc. El efecto de corrección de esta lámpara para los cabezales Raman empleados con el espectrómetro de doble pista se muestra en la Figura 13.

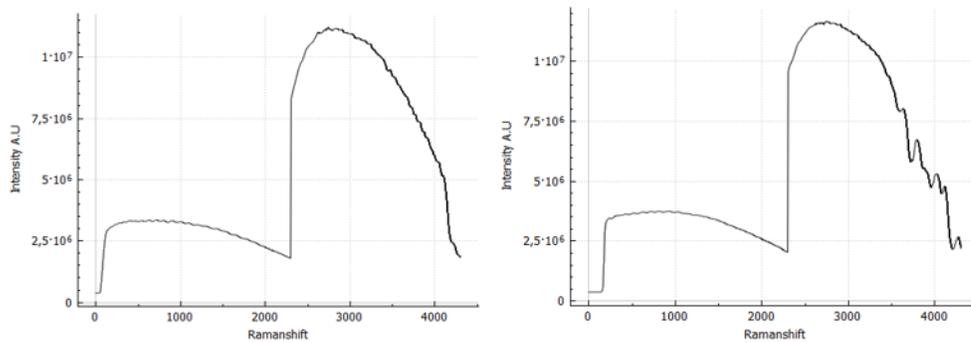


Figura 13: De izquierda a derecha, corrección para los cabezales BWTECK-2011 (“RAD-2”) y BWTECK-2008 (“i-Raman”). Puede apreciarse la diferencia de comportamiento de ambos espectrómetros frente a la misma señal de entrada.

2.1.4. Ciclohexano

Para la calibración de los instrumentos también se ha empleado una muestra de ciclohexano certificada por la ASTM (American Society for Testing and Materials), comercializado por la marca Kaiser Optical Systems, Inc como parte del set Raman Calibration Accesory. El ciclohexano es uno de los ocho materiales seleccionados por la ASTM como estándar para la calibración de los desplazamientos Raman a causa de la estabilidad de sus picos (10) y (11). Su picos Raman característicos pueden observarse en Figura 14..

384.1 ± 0.78	6	2
426.3 ± 0.41	6	3
801.3 ± 0.96	6	95
1028.3 ± 0.45	6	15
1157.6 ± 0.94	6	6
1266.4 ± 0.58	6	14
1444.4 ± 0.30	6	12
2664.4 ± 0.42	6	8
2852.9 ± 0.32	6	100
2923.8 ± 0.36	6	58
2938.3 ± 0.51	6	67

Figura 14: Bandas Raman características del ciclohexano (21)

2.2. CALIBRACIÓN DE ESPECTRÓMETROS Y ESPECTROS

2.2.1. Calibración del espectrómetro de doble pista

La calibración del espectrómetro es un proceso que merece especial cuidado puesto que la precisión en la posición de los picos Raman dependerá de ella. La información espectral es recogida por la cámara del espectrómetro en píxeles. Inicialmente el espectro es una representación bidimensional de intensidad frente a píxel. La información espectral será útil si se conoce a qué desplazamiento Raman corresponde cada píxel.

Buscando conocer la equivalencia existente entre píxel y desplazamiento Raman se mide el espectro luminoso de las lámparas de calibración (Hg-Ar y Ne), así como la señal Raman del ciclohexano.

Una vez adquiridas las señales de las lámparas es necesario transformar sus posiciones en desplazamientos Raman (número de onda con respecto al número de onda de la excitación) mediante la sencilla ecuación 17.

$$rshift = \frac{1}{\lambda_{laser}} - \frac{1}{\lambda_{lampara}} \quad (17)$$

Conviene notar que en el caso de las lámparas no es necesaria una fuente de excitación láser, pues ya emiten por sí mismas, pero sí que debe hacerse la conversión a desplazamiento Raman en base a la longitud de onda de excitación que utilizará el equipo.

Laser(nm) 532			Laser(nm) 532		
Track 1	l(nm)	raman(cm-1)	Track 2	l(nm)	raman(cm-1)
	597,553	2062,075244			
	603	2213,244554	1306,768	585,249	1710,248207
94,631	607,434	2334,298592	1383,986	588,189	1795,654476
135,742	609,616	2393,223551	1554,035	594,483	1975,653604
225,689	614,306	2518,460284	1640	597,553	2062,075244
265,718	616,359	2572,681649	1797,98	603	2213,244554
369,867	621,728	2712,788456	1931,037	607,434	2334,298592
466,734	626,649	2839,095796	370,936	546,074	484,456085
542,908	630,479	2936,03597	1095,006	576,96	1464,76841
602,118	633,443	3010,252396	1147,999	579,066	1527,803822
700,938	638,299	3130,353492	62,939	532	0
740,307	640,225	3177,483715	587,767	555,6885355	801,3
959,129	650,653	3427,817207	751,998	562,7876114	1028,3
1015,934	653,288	3489,807901	849,963	566,9129484	1157,6
1160,825	659,895	3643,066478	934,092	570,4313765	1266,4
			1077,132	576,2827664	1444,4

Lámparas:
 Neón
 Mercurio-Argón

Espectro Raman:
 Ciclohexano

Figura 15: Datos píxel-longitud de onda-desplazamiento Raman de las lámparas y espectros empleados en la calibración.

Con dichos datos adquiridos (ver Figura 15.) y debidamente transformados se crea la curva de calibración, que se utilizará para obtener un polinomio que relacione el píxel con la longitud de onda y el desplazamiento Raman. En el caso que nos ocupa, esta curva se ajustó mediante un polinomio de tercer orden. En las Figuras 16. y 17. se muestran dichas curvas para las dos pistas del espectrómetro.

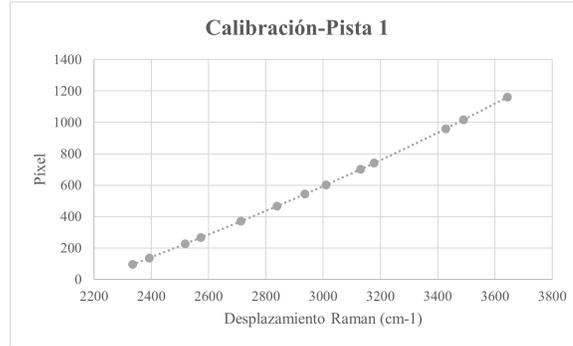


Figura 16: Polinomios de calibración para la la primera pista del espectrómetro.

Cuyo polinomio de calibración se muestra en la siguiente ecuación (ecuación 18):

$$D_{R1}[cm^{-1}] = (2,41e-8 \pm 0,13e-8)p^3 - (11,7e-5 \pm 1,2e-5)p^2 + (0,85 \pm 0,03)p - (1570 \pm 30) \quad (18)$$

donde se incluye el error accidental asociado a cada coeficiente. D_{R1} denota el desplazamiento Raman en cm^{-1} asociado a cada píxel p para la primera pista del espectrómetro. El coeficiente de correlación de este polinomio con los datos obtenidos es $r_1 = 0,99999967$

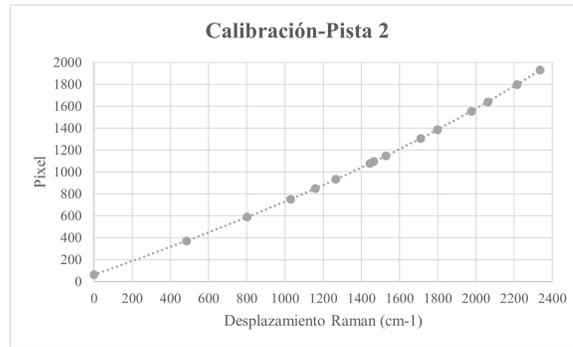


Figura 17: Polinomios de calibración para la la segunda pista del espectrómetro.

El polinomio de calibración asociado a esta segunda pista se muestra en la siguiente ecuación (ecuación19):

$$D_{R2}[cm^{-1}] = (2,44e-8 \pm 0,12e-8)p^3 + (1,7e-5 \pm 0,4e-5)p^2 + (0,627 \pm 0,004)p - (62,2 \pm 1,2) \quad (19)$$

con coeficiente de correlación

$$r_2 = 0,99999544$$

Nota sobre la calibración:

Como puede apreciarse en la imagen superior, el parámetro elegido como variable independiente en la calibración es el desplazamiento Raman (con valores enteros o con incrementos constantes), obteniendo la correlación del desplazamiento Raman con respecto al pixel correspondiente. Esta selección se fundamenta en que los equipos antiguos estaban preparados para trabajar con vectores equiespaciados de desplazamientos Raman. De este modo, es necesario conocer con qué píxel se corresponden los distintos desplazamientos de dicho vector. Para aprovechar al máximo en este proceso la información contenida en el espectro, se interpola la intensidad correspondiente a los dos píxeles más cercanos al valor proporcionado por el polinomio de calibración. El paso en desplazamiento Raman se efectuaba con medias unidades para lograr una mayor precisión.

En cierto momento de revisión del calibrado se decidió realizar la calibración convirtiendo cada uno de los píxeles en desplazamiento Raman, de forma que no se perdiera resolución espectral por efecto de la interpolación necesaria al hacer la calibración a valores enteros de desplazamiento Raman.

2.2.2. Calibración de los espectros adquiridos y creación de SpectrumCalibrator

Para llevar a cabo la calibración de los espectros adquiridos con el espectrómetro de doble pista tres procesos son posibles:

- Calibración con programas de ANDOR basis con comandos internos que ejecuta software de adquisición.
- Calibración con MATLAB.
- Calibración empleando SpectrumCalibrator, una nueva aplicación que, usando código de MATLAB, permite calibrar espectros sin la necesidad de programar.

Inicialmente, el método utilizado era el segundo, haciendo uso de scripts previamente empleados por otros miembros de ERICA. Sin embargo, otra operaria pasó a adquirir espectros con el mismo equipo, pero sin poseer conocimiento de MATLAB. A simple vista podría parecer que la mejor solución es que la calibración pasara a efectuarse directamente con el programa interno del ANDOR Newton. Este fue el proceso seguido al comienzo, pero debido a una comparación se vio que ambas calibraciones no eran equivalentes (véase Figura 18.) . La calibración interna de ANDOR Newton deformaba los espectros desplazando sus picos Raman característicos, tal y como puede verse en la Figura 18.. Según *S.K. Sharma* (12) las características espectrales de la microclina empleada en la comparación han de ser las obtenidas con la calibración de MATLAB.

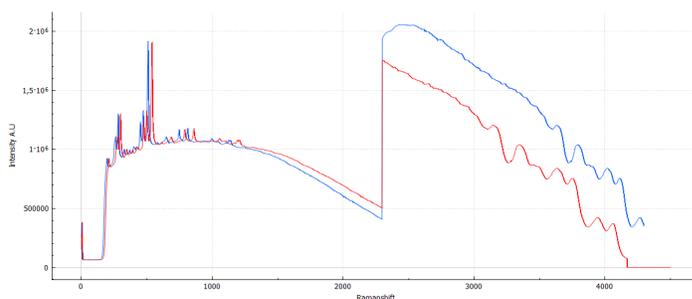


Figura 18: Calibración del mismo espectro de microclina usando MATLAB (en azul) y ANDOR (en rojo). El pico principal de la muestra se encuentra en 510cm^{-1} en la calibración de MATLAB y en 539cm^{-1} en la de ANDOR. La bibliografía sitúa este pico en 515m^{-1} .

Ante esta situación se probaron dos soluciones: un primer intento fue explorar el código interno del programa del software de la cámara ANDOR Newton (ver Figura 19.).

Se probó a redefinir el tamaño de los vectores y a establecer un paso en los bucles de media unidad en lugar de una entera sin lograr mejora. Finalmente, se probó a realizar la calibración al revés, buscando la equivalencia exacta de cada píxel con un cierto desplazamiento Raman. Al efectuar este cambio en ANDOR se descubrió que en la zona central del espectro solapan las dos pistas del espectrómetro. Esta nueva forma de calibrar también proporcionaba espectros ligeramente deformados.

```

update(0)

rem Calibra sobre la primera pista (Track 2) que es la de desplazamiento menor
rem Utiliza el Track 1 hasta los 2300 cm-1
i=1
while(i<2300)
rem Calibración del desplazamiento raman a pixel de la primera pista como
rem polinomio de 3º grado

n= 2.4403973*10^(-8)*i*i*i + 1.6922772*10^(-5)*i*i + 6.2734039*10^(-1)*i + 6.2162675*10^(1)

v1=mod(n,1)
j=n-v1
s1=#0_sig(j,3)
s2=#0_sig(j+1,3)
rem asigna al valor de desplazamiento raman la intensidad por interpolación
#1[i]=s1+(s2-s1)*v1
i=i+1
wend
while(i<4501)
rem Calibración del desplazamiento raman a pixel de la segunda pista como
rem polinomio de 3º grado

n= 2.4125808*10^(-8)*i*i*i - 1.1658647*10^(-4)*i*i + 8.5487416*10^(-1)*i - 1.5728463*10^(3)

v1=mod(n,1)
j=n-v1
s1=#0_sig(j,1)
s2=#0_sig(j+1,1)
rem asigna al valor de desplazamiento raman la intensidad por interpolación
#1[i]=s1+(s2-s1)*v1
i=i+1
wend
update(1)

```

Figura 19: Fragmento de código del programa de calibración de ANDOR.

Finalmente, se optó por crear una aplicación que permitía calibrar espectros desde cualquier ordenador independientemente de que el usuario posea o no MATLAB. Esta nueva aplicación incluye la ventaja de aprovechar al máximo la resolución espectral del instrumento, buscando el desplazamiento Raman que equivale a cada píxel, de modo que el paso en píxeles se efectúa en unidades y el vector de desplazamiento Raman no posee unidades equiespaciadas. Con los actuales equipos y softwares utilizados, la no equidistancia del desplazamiento Raman no supone un problema. Además, las herramientas modernas de tratamiento espectral permiten realizar la interpolación de los datos de una forma fácil para obtener espectros con una mayor densidad de puntos. La nueva aplicación ha sido llamada *SpectrumCalibrator* en honor a la función que realiza.

SpectrumCalibrator

SpectrumCalibrator es una aplicación desarrollada como parte de este trabajo que sirve para el calibrado de espectros Raman adquiridos con espectrómetros de doble pista, si bien está parametrizado para el sistema Kayser utilizado en el presente estudio. Ciertos parámetros internos como la cantidad de píxeles de la CCD, el punto de separación de las dos pistas y el orden del polinomio de calibración son específicos para el instrumento utilizado y deberían ser modificados en el código fuente de la aplicación para poder alterarlos. El icono de escritorio de SpectrumCalibrator se muestra en la Figura 20..



Figura 20: Icono característico de SpectrumCalibrator.

La aplicación consiste en un ejecutable de MATLAB, compilado con la herramienta deploytool. Consta de una sencilla interfaz que muestra al usuario el polinomio de calibración en uso (Figura 21.) y le permite elegir si quiere o no modificar los coeficientes de dicho polinomio (Figura 22.), que se actualizan por defecto cada vez que son corregidos.

Una vez determinado el polinomio de calibración el usuario debe seleccionar los espectros que desea calibrar y el programa le devuelve tres archivos con formato .asc (formato en el cual guarda los espectros el programa ANDOR Newton).

Los archivos creados contienen:

- *calibrated-nombre-del-espectro.asc* Espectro calibrado sin solapamiento entre pistas. Se toma la información Raman de la pista número 2 (pista de pequeños desplazamientos) hasta 2300cm⁻¹ y la de la pista 1 (de grandes desplazamientos) a partir de dicho valor.
- *calibrated-track1-nombre-del-espectro.asc* Espectro que contiene la información completa de la pista número 1, que se corresponde con grandes desplazamientos Raman.
- *calibrated-track2-nombre-del-espectro.asc* Espectro que contiene la información completa de la pista número 2, que se corresponde con pequeños desplazamientos Raman.

Los archivos de las dos pistas se guardan por separado para no perder nada de la información contenida en los espectros originales.

Tras calibrar los espectros, la aplicación devuelve una gráfica con la representación de todos ellos y una leyenda con sus nombres (Figura 23.), de modo que el experimentador puede observar fácilmente las características de los espectros que ha adquirido, así como comparar unos con otros.

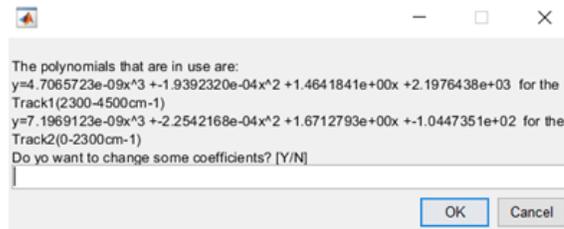


Figura 21: Interfaz de SpectrumCalibrator que muestra los polinomios en uso.

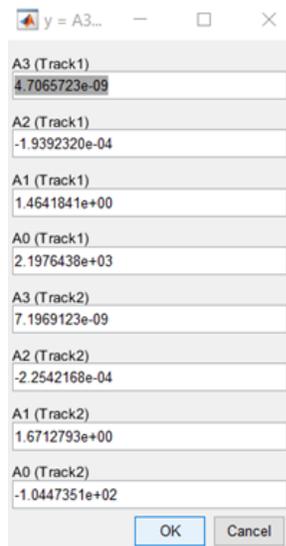


Figura 22: Interfaz de SpectrumCalibrator para la modificación de los coeficientes de calibración.

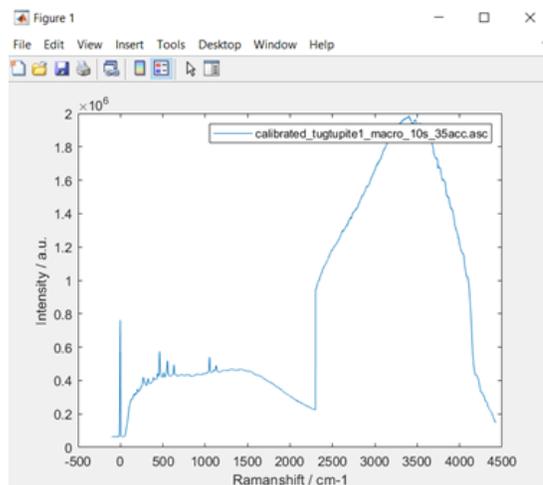


Figura 23: Interfaz de SpectrumCalibrator que muestra los espectros calibrados.

2.2.3. Calibración del RLS Sim

El procedimiento de calibración del simulador del instrumento RLS es muy similar al explicado previamente para el espectrómetro de doble pista. La diferencia principal es que en este caso también es necesario caracterizar la posición píxel en la que aparece el pico de emisión del láser, así como su longitud de onda exacta, ya que en este caso no estaba definida.

En el caso de la calibración del espectrómetro de doble pista la posición en píxel de los distintos picos era proporcionada por el programa ANDOR Newton. Los espectros adquiridos con este espectrómetro eran espectros con una línea de base muy pequeña fácilmente despreciable y con picos estrechos.

Los espectros de RLS-Simulator, sin embargo, son espectros que guardan la antigua calibración y por tanto han de transformarse en espectros de píxel-intensidad. Para ello se sustituye el vector de desplazamiento Raman por el vector de píxeles correspondiente con su CCD (2068 píxeles equiespaciados). Estos espectros, a su vez, cuentan con picos anchos y con un cierto crecimiento en su línea de base. Para que la calibración fuera lo más precisa posible era necesario procesar los espectros eliminando su línea de base con una curva de corrección muy noble. Dicha corrección de línea de base solo ha sido aplicada al espectro de ciclohexano.

Eliminación de la línea de base en la calibración.

La línea de base característica del simulador del instrumento RLS posee una zona ancha en la parte central que disminuye progresivamente hacia los extremos del espectro. Esto hace que los picos de la zona baja del espectro estén ligeramente desplazados hacia la derecha, debido a la mayor contribución de la intensidad de la línea de base en esta zona del pico, así como los picos de la zona alta del espectro están ligeramente desplazados a la izquierda.

Una vez corregida la línea de base se somete a los espectros de calibración a un ajuste de bandas para determinar con exactitud la posición del pico Raman. Se opta por este procedimiento debido a que el tamaño de píxel es del orden de magnitud de la resolución espectral. Lo más habitual es que el máximo del pico (en píxel) no siempre coincida con la máxima posición de intensidad de emisión. Por tanto, se hace este ajuste en un procedimiento de cálculo de la posición conocido como procesamiento sub-píxel. En la Figura 24. puede observarse este ajuste para las bandas empleadas en el caso del ciclohexano.

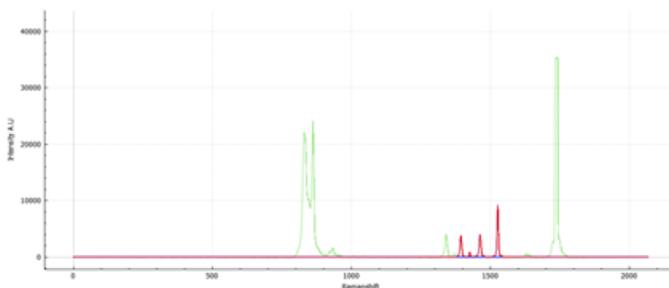


Figura 24: El ajuste de bandas solo se lleva a cabo en los picos necesarios para la calibración.

A la par, debido a problemas de envejecimiento del láser empleado, su longitud de onda se ha visto modificada con el paso de los años. Para lograr determinar con precisión la nueva longitud de onda se emplea la información Raman contenida en el espectro de ciclohexano, llegando a obtener la nueva longitud de onda del láser mediante un proceso iterativo descrito a continuación.

Con la posición de píxel medida para las cuatro bandas características del ciclohexano (Tabla 2.) calculamos la longitud de onda correspondiente a cada una de ellas. Para ello se emplea un polinomio de calibración calculado con las posiciones de las lámparas y el láser (Figura 25.).

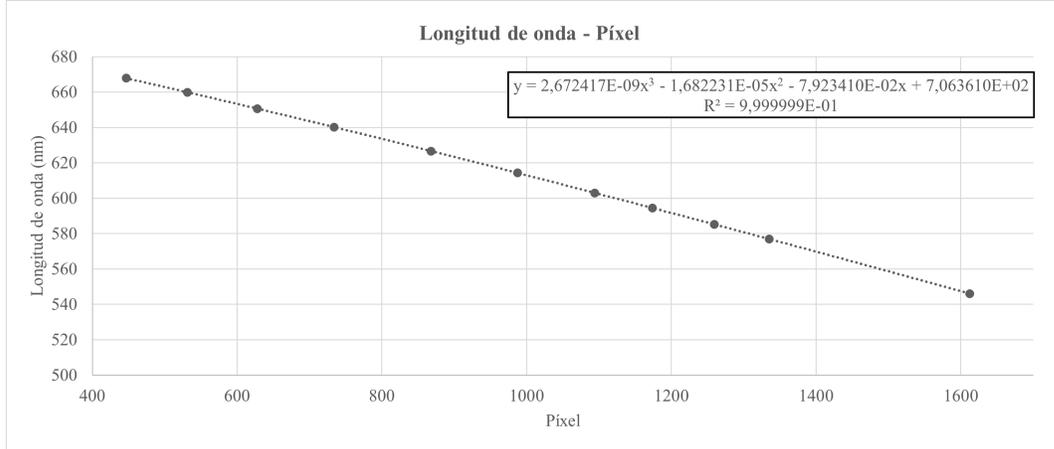


Figura 25: Polinomio de orden 3 para la conversión de píxel a longitud de onda de los picos de ciclohexano. La curva ha sido creada con puntos equiespaciados de los espectros de emisión de las lámparas y el láser.

La posición Raman característica de estos picos del ciclohexano es conocida y mediante un ajuste por mínimos cuadrados (Figura 26.) a la ecuación 20:

$$rshift[cm^{-1}] = \left(\frac{1}{\lambda_{laser}} - \frac{1}{wavelength[nm]} \right) \cdot 10^7 \quad (20)$$

podemos calcular la longitud de onda del láser. En lugar de quedarnos con el primer valor que resulta de este ajuste, nos damos cuenta de que el polinomio de calibración píxel-longitud de onda se modifica si modificamos λ_{laser} de modo que el proceso se lleva a cabo recursivamente hasta que la longitud de onda del láser toma un valor estable con las nuevas iteraciones, de forma que se obtiene el polinomio óptimo de calibración que minimiza el error del ajuste.

Píxel	λ calculada (nm)	Desplazamiento Raman cm^{-1}	$\frac{1}{\lambda}(nm^{-1})$
1526,07	555,7388133	801,03	0,0017799406
1462,4	562,8593677	1028,3	0,00176643
1425,2	555,7388133	1157,5	0,001763672
1393,3	555,7388133	1266,4	0,001752739

Tabla 2: Datos obtenidos mediante el polinomio de calibración longitud de onda – píxel.

Siguiendo este procedimiento se obtiene una longitud de onda para el láser de $\lambda_{laser} = 531,970 \pm 0,018nm$.

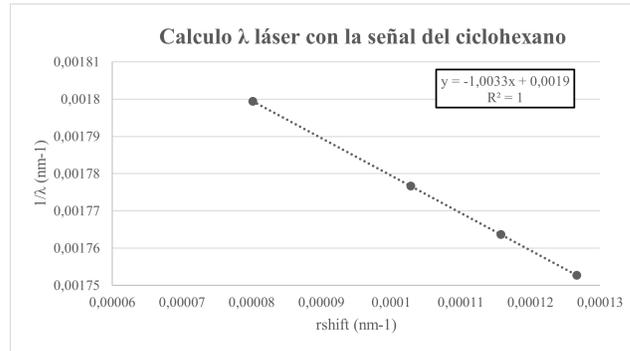


Figura 26: Ajuste por mínimos cuadrados para el cálculo la longitud de onda del láser.

	λ (nm)	Píxel	Desplazamiento Raman (cm^{-1})
Laser	531.97	1735.84	0
Hg	546.074	1612.32	485.5161256
Ciclohexano	555.6558045	1526.07	801.3
Ciclohexano	566.8788818	1425.2	1157.6
Hg	576.96	1334.97	1465.828451
Ne	585.24879	1259.25	1711.302116
Ne	594.483	1173.85	1976.713644
Ne	603	1094.2	2214.304595
Ne	614.306	987.4	2519.520324
Ne	626.649	868	2840.155836
Ne	640.225	733.8	3178.543755
Ne	650.653	627.9	3428.877247
Ne	659.895	531.5	3644.126518
Ne	667.828	446.85	3824.137082

Tabla 3: Datos seleccionados para la calibración del instrumento RLS Sim.

Una vez calculada la longitud de onda del láser con precisión se usa su valor para la calibración final del instrumento. De todos los picos medidos para las distintas lámparas y el ciclohexano se seleccionan 14 de ellos (véase Tabla 3.) de manera aproximadamente equiespaciada para que el peso del polinomio de calibración sea similar en todas las regiones del espectro.

Finalmente, el polinomio de calibración obtenido se muestra en la ecuación 21:

$$D_R[cm^{-1}] = (-9,8e - 8 \pm 1,1e - 8)p^3 - (4,3e - 4 \pm 0,4e - 4)p^2 - (1,64 \pm 0,04)p + (4653 \pm 14) \quad (21)$$

Donde se mantiene la notación empleada en la calibración del espectrómetro de doble pista.

En la Figura 27 se observa que los puntos seleccionados para la calibración final se encuentran equiespaciados a lo largo del espectro.

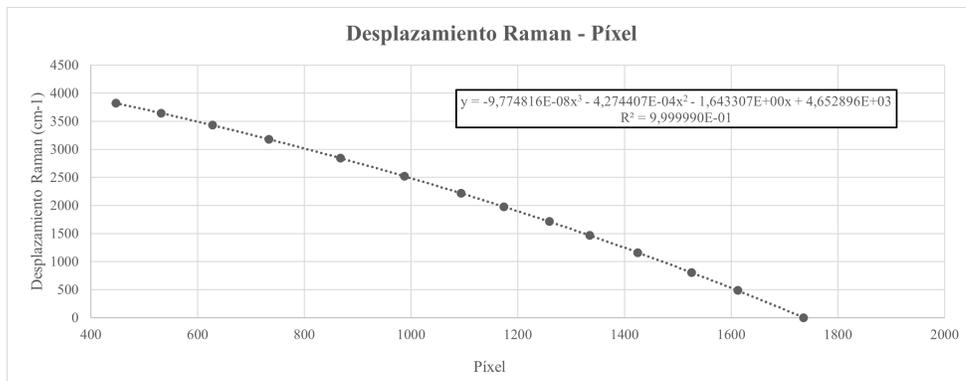


Figura 27: Polinomio de calibración del instrumento RLS Sim

Una vez obtenido este polinomio se puede calcular el vector de desplazamiento Raman a partir del vector de píxeles característico del espectrómetro. Con la nueva calibración se procede a la toma de datos.

2.3. ADQUISICIÓN DE ESPECTROS PARA LA BASE DE DATOS ADAMM

La base de datos ADAMM (14) pretende dar soporte a las misiones marcianas que cuentan con espectrómetros Raman. Para ello, se pretende crear una amplia colección de espectros adquiridos con instrumentos similares a los que se encuentran a bordo del rover Perseverance de la misión Mars2020 y Rosalind Franklin de la misión ExoMars2022. La base de datos, actualmente en desarrollo, proporcionará datos adquiridos con diversas técnicas, entre las que se incluyen difracción de rayos-X y espectroscopia infrarroja.

Los minerales han sido seleccionados atendiendo a lo que se espera encontrar en Jezero Crater y Oxia Planum:

- Piroxenos y piroxenoides
- Feldespatos y feldespatooides
- Olivinos
- Anfíboles
- Epidotas
- Granates
- Sílice
- Óxidos e hidróxidos de Fe-Mn-Ti
- Filosilicatos
- Zeolitas
- Carbonatos
- Sulfatos

Para la adquisición de espectros destinados a formar parte de la base de datos ADAMM se ha empleado el espectrómetro Raman de doble pista. Las muestras adquiridas han sido diversos feldespatos, feldespatooides y piroxenos de interés para la caracterización marciana. Los ejemplares empleados eran muestras en bloque.

Antes de comenzar el proceso de adquisición de espectros, es necesario, en primer lugar, calibrar el espectrómetro como ya se ha explicado en apartados anteriores; después, estudiar los parámetros óptimos de adquisición para cada una de las muestras y puntos de adquisición; adquirir los espectros y darks con el mismo tiempo de integración; calibrar los espectros, y finalmente correlar con la información disponible en la literatura las señales obtenidas en dichos espectros.

Una vez verificada la correcta adquisición del espectro se procede al acondicionamiento del mismo mediante la obtención del espectro promedio de todas las acumulaciones (división por el número de acumulaciones), resta del dark, división entre la lámpara de calibración del instrumento y eliminación de la línea de base. Una vez completado este proceso, el espectro se incorpora a la base de datos ADAMM, convirtiéndose en un estándar de referencia.

2.3.1. Muestras adquiridas

En este apartado se recogen las muestras adquiridas durante la realización de este trabajo, así como su clasificación y características más importantes.

Feldespatos Los feldespatos son un tipo de mineral de silicio: silicatos, concretamente tectosilicatos, en los que los tetraedros de sílice cristalizan compartiendo sus cuatro vértices dando lugar a una estructura tridimensional. Los feldespatos tienen por fórmula general $(Si, Al)_4O_8(Na, K, Ca)$, donde el Silicio y el Aluminio intercambian sus posiciones en función de la abundancia de dichos elementos en el ambiente de formación. La sustitución puede variar entre un 25 % y un 50 %. Los feldespatos se clasifican en dos grandes grupos: los feldespatos alcalinos y las plagioclasas. Esta clasificación puede verse en la Figura 28..

Los feldespatos analizados son:

- | | | |
|--------------------|--------------|---------------|
| ■ albita | ■ microclina | ■ oligoclasa |
| ■ ortosa/ortoclasa | ■ amazonita | ■ andesina |
| ■ anortoclasa | ■ albita | ■ labradorita |
| ■ sanidina | ■ anortita | ■ celsiana |

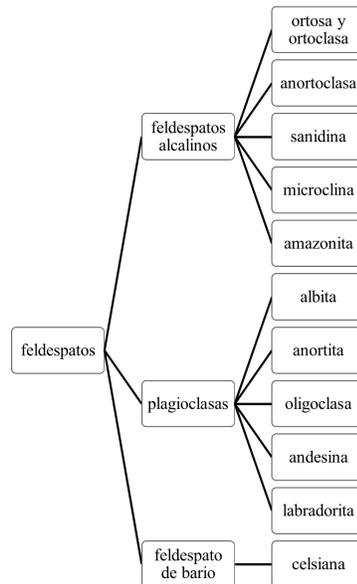


Figura 28: Clasificación de los distintos feldespatos analizados en función de su grupo mineral.

Muchas de las muestras son proporciones intermedias entre extremos de composición precisa. La existencia de estas fases compositivas intermedias da lugar a lo que geológicamente se denomina solución sólida. Un ejemplo de la principal solución sólida para los feldespatos puede observarse en la Figura 29..

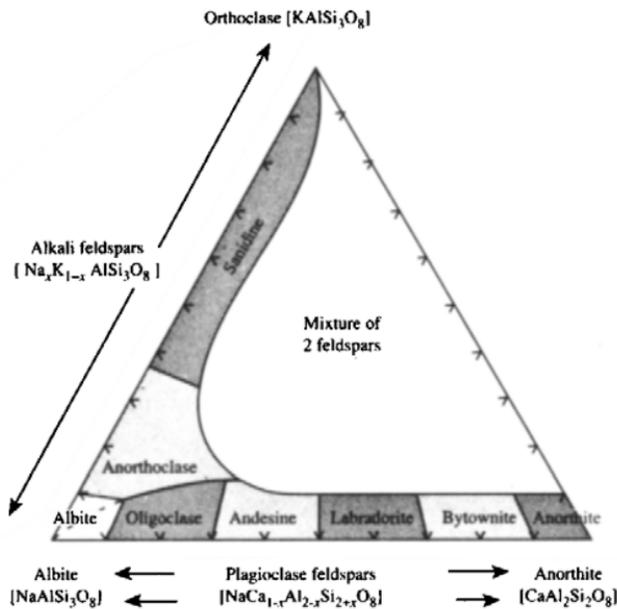


Figura 29: Las rocas volcánicas de Marte presentan una amplia variedad de elementos de soluciones sólidas de feldspatos (13) alcalinos y plagioclasas. En el esquema se muestran las soluciones sólidas más comunes de los feldspatos.

Feldespatoideos Los feldespatoideos se forman en rocas volcánicas con déficit de sílice y abundancia de álcalis. Algunos feldespatoideos poseen hidratación, lo que produce una ancha señal Raman en la zona de altos desplazamientos. Los feldespatoideos analizados en este trabajo han sido los siguientes:

- Leucita
- Noseana
- Cancrinita
- Nefelina
- Haiüyna
- Sodalita
- Lazurita
- Analcima

Piroxenos Los piroxenos son minerales de silicio, concretamente inosilicatos con estructura de cadena de tetraedros de SiO_4 . Los piroxenos poseen simetría rómbica o monoclinica. Se clasifican en cuatro grupos:

- Piroxenos ferromagnesianos
- Piroxenos sódicos
- Piroxenos cálcicos
- Piroxenos calcosódicos

En este caso se han tomados espectros de distintas muestras de un piroxeno cálcico denominado augita $(Ca, Mg, Fe^{2+})_2[Si, Al_2O_6]$, que ha sido tomado como mineral de referencia para este grupo.

2.3.2. Parámetros de adquisición y toma de muestras

Los parámetros con los que se han adquirido los diversos espectros se muestran en las siguientes tablas (Tabla 4. y Tabla 5.):

Tipo de muestra	Muestra	Ti (s)	Na (acc)	LASER pow (mW)
Feldespatos	Albite 3	130	10	22.7
Feldespatos	Labradorite 2	160	5	22.7
Feldespatos	Sanidine 1	120	10	22.7
Feldespatos	Sanidine 1	120	21	22.7
Feldespatos	Andesine 1	40	12	22.7
Feldespatos	Anorthoclase 1	90	12	22.7
Feldespatos	Orthoclase 1	15	30	48
Feldespatos	Orthoclase 3	60	30	48
Feldespatos	Orthoclase 4	22	40	48
Feldespatos	Albite 2	3.5	100	48
Feldespatos	Oligoclase 3	21	30	48
Feldespatos	Ortosa 2	3.5	150	48
Feldespatos	Oligoclase 2	9	30	48
Feldespatos	Orthoclase 1	10	30	48
Feldespatos	Orthoclase 1	5	30	48
Feldespatos	Orthoclase 3	36	30	48
Feldespatos	Albite 1	23	30	48
Feldespatos	Amazonite 1	40	12	48
Feldespatos	Microcline 1	13	35	48
Feldespatos	Amazonite 2	45	20	48
Feldespatos	Microcline 2	43	20	48
Feldespatos	Oligoclase 1	17	25	48
Feldespatos	Ortosa 6	2.7	30	48
Feldespatos	Albite 4	27	30	48
Feldespatos	Anorthite 1	75	12	48
Feldespatos	Microcline 3	2.6	35	48
Feldespatos	Microcline 3	23	16	48
Feldespatos	Celsian 1	50	18	48
Feldespatos	Celsian 1	15	10	48
Feldespatos	Sanidine 2	20	40	48
Feldespatos	Andesine 1	120	20	48
Feldespatos	Sanidine 1	65	20	48
Feldespatos	Oligoclase 2	9.2	30	48
Feldespatos	Oligoclase 3	34	21	48
Feldespatos	Orthoclase 1	3.2	40	48
Feldespatos	Ortosa 2	2.4	40	48
Feldespatos	Albite 2	1.7	30	48
Feldespatos	Albite 3	26	15	48
Feldespatos	Labradorite 2	103	20	48
Feldespatos	Anorthoclase 1	6	30	48

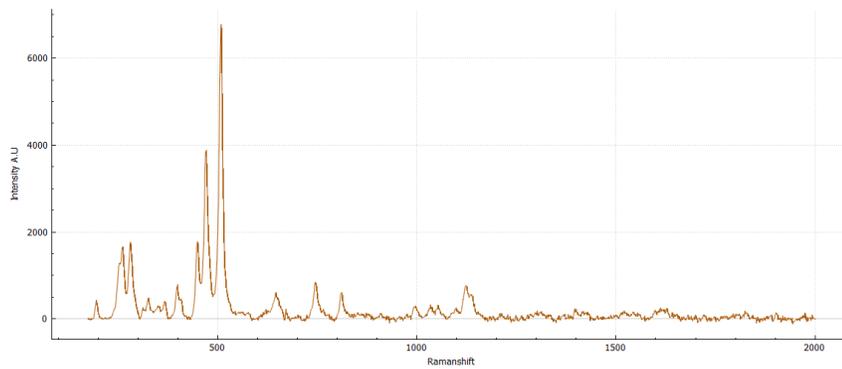
Tabla 4: Parámetros de adquisición de los espectros para la base de datos ADAMM. Parte 1.

Tipo de muestra	Muestra	Ti (s)	Na (acc)	LASER pow (mW)
Feldespatos	Orthoclase 4	4	30	48
Feldespatos	Albite 1	22	20	48
Feldespatos	Orthoclase 3	21	20	48
Feldespatos	Ortosa 1	36	25	48
Feldespatos	Sanidine 2	30	15	48
Feldespatos	Amazonite 1	37	20	48
Feldespatos	Labradorite 1	100	15	48
Feldespatos	Labradorite 2	72	18	48
Feldespatos	Oligoclase 2	20	27	48
Feldespatos	Oligoclase 1	15	40	48
Feldespatos	Albite 4	30	30	48
Feldespatos	Celsian 1	20	30	48
Feldespatos	Amazonite 2	50	24	48
Feldespatos	Anorthite 1	100	12	48
Feldespatos	Microclina 3	29	26	48
Feldespatos	Ortosa 3	18	40	48
Feldespatos	Microcline 2	25	30	48
Feldespatos	Orthoclase 2	24	30	48
Feldespatos	Ortosa 5	27	30	48
Feldespatos	Ortosa 4	50	20	48
Feldespatos	Microcline 1	13	34	30.7
Feldespatoides	Analcime 1	60	20	30.7
Feldespatoides	Petalite 2	30	20	30.7
Feldespatoides	Analcime 2	40	18	30.7
Feldespatoides	Nepheline 1	60	16	40.8
Feldespatoides	Cancrinite 1	50	18	40.8
Feldespatoides	Sodalite 2	15	20	40.8
Feldespatoides	Sodalite 2	10	25	40.8
Feldespatoides	Sodalite 1	40	20	40.8
Feldespatoides	Sodalite 1	3	30	40.8
Feldespatoides	Hauyne 1	15	20	40.8
Feldespatoides	Analcime 3	21	25	40.8
Feldespatoides	Petalite 1	22	36	40.8
Feldespatoides	Tugtupite 1	15.5	45	40.8
Feldespatoides	Leucite 1	24	30	40.8
Feldespatoides	Lazurite 1	12	35	40.8
Feldespatoides	Cancrinite 1	16	20	40.8

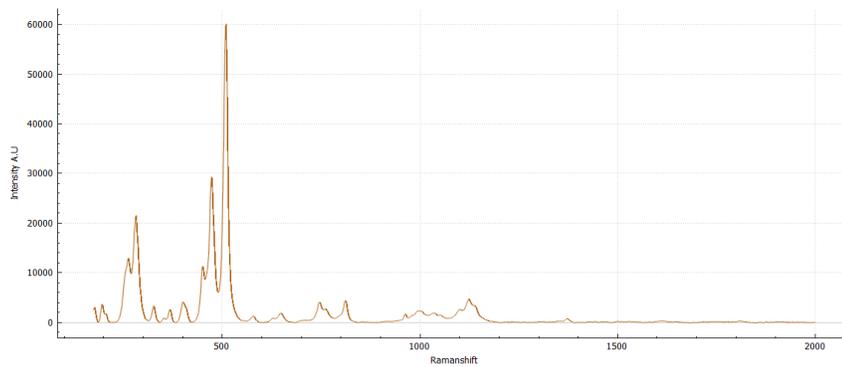
Tabla 5: Parámetros de adquisición para los espectros de la base de datos ADAMM. Parte 2.

El tiempo de integración depende del comportamiento de la muestra y de la potencia del láser, la cual se ha visto modificada por el cambio de fuentes y cabezales realizado. Todos los espectros han sido adquiridos en condiciones de máxima oscuridad para evitar la contribución de luz ambiental, especialmente, aquella procedente de las lámparas fluorescentes. Para proteger la vista del experimentador de la radiación láser, durante la manipulación de las muestras se han empleado gafas protectoras.

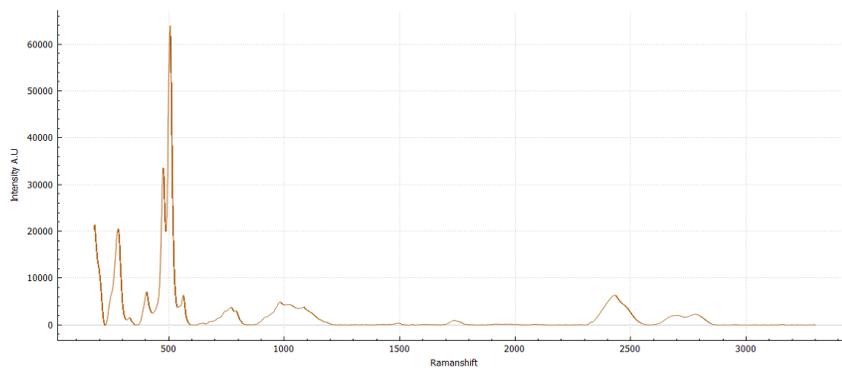
A continuación se muestra una selección de espectros significativos de entre los adquiridos para este estudio: ortoclasa (Figura 30a.), como ejemplo de feldespato alcalino y albita (Figura 30b.) y labradorita(Figura 30c.), como plagioclasas.



(a) Espectro de ortoclasa 1 adquirido para la base de datos ADAMM



(b) Espectro de albita 1 adquirido para la base de datos ADAMM



(c) Espectro de labradorita 1 adquirido para la base de datos ADAMM

Figura 30: Muestra de espectros adquiridos para la base de datos ADAMM.

2.4. ADQUISICIÓN DE ESPECTROS DE MUESTRAS PROBLEMA

Para la verificación de los algoritmos desarrollados para la identificación automática es necesario realizar la adquisición de espectros de muestras problema, que se utilizarán para dicho fin. Esta adquisición se ha llevado a cabo con el simulador del instrumento RLS, con el objetivo de reproducir de la forma más precisa posible las características de los espectros que llegarán desde Marte, así como de evitar un posible sobre-entrenamiento de los modelos al utilizar la misma equipación para definir los espectros de la base de datos y los espectros problema. Con el fin de replicar el funcionamiento del instrumento RLS durante su adquisición de muestras marcianas, las muestras empleadas en esta parte del estudio son muestras en polvo que se analizan en diferentes puntos. La adquisición de estos espectros se ha llevado a cabo automáticamente gracias a las rutinas que permite ejecutar el RLS Sim (15).

Tal y como quedó explicado en 2.1.2, los únicos parámetros que hubo que definir para llevar a cabo la adquisición de los espectros se muestran en la Tabla 6..

Parámetro	Valor
Potencia láser	70 %
Puntos por muestra	6
N° acumulaciones	40
N_{av}	1

Tabla 6: Parámetros de adquisición de los espectros de las muestras problema.

2.4.1. Muestras adquiridas

Los materiales seleccionados como muestras problema son minerales puros y mezclas binarias. Estas muestras están compuestas de dos tipos de silicatos: nesosilicatos y filosilicatos. Los minerales que las forman se describen a continuación:

Olivino El olivino es un nesosilicato, es decir, un mineral formado por tetraedros aislados de $(SiO_4)^{4-}$ unidos por enlaces iónicos a cationes. Concretamente, la variedad de olivino estudiada es la fosterita, Mg_2SiO_4 , que posee el color verde oliva característico que da nombre al grupo.

Serpentina La serpentina es un grupo dentro de los filosilicatos de fórmula química $Mg_3[Si_2O_5](OH)_4$. Entre sus variedades destacan la lizardita, el crisotilo y la antigorita.

Vermiculita La vermiculita es un filosilicato de composición $(Al, Si)_4O_{10}(Mg, Fe, Al)_3(OH)_2Mg_{0,3}4H_2O$. Las variaciones de composición dentro de la fórmula dan lugar a varios tipos de vermiculita.

Caolinita La caolinita es un filosilicato de fórmula química $Al_2[Si_2O_5](OH)_4$ de color blanco y composición química muy constante.

Las muestras puras y mezclas binarias empleadas en la mejora del método son las siguientes:

- Olivino 100 %
- Serpentina 100 %
- Vermiculita 100 %
- Caolinita 100 %
- Vermiculita-Caolinita 50 %
- Serpentina-Olivino 47,5 %

Las muestras han de ser colocadas en la platina del simulador con una espátula. Para facilitar el enfoque automático sobre la muestra esta se aplanando usando una pieza metálica tal y como se muestra en la Figura 31.



Figura 31: Platina con muestras de entrenamiento y verificación. El aplanado de la muestra facilita el enfoque automático.

2.5. CONCLUSIONES DEL TRABAJO EXPERIMENTAL

Podría parecer que la espectroscopia Raman se limita a la interacción que sufre la luz con la material. Sin embargo, al ponerla en práctica, tal y como sucede con otras técnicas experimentales, se ve que el proceso es mucho más largo y complejo. Partiendo de la calibración de los instrumentos, puede apreciarse que el procedimiento es meticuloso y que los resultados obtenidos han de ser precisos. De lo contrario, se desaprovechará la información espectral.

Prosiguiendo con la adquisición de los espectros, es necesario destacar que cada muestra puede presentar un comportamiento muy dispar a las anteriores, ya que la eficiencia de emisión es muy variable debido a las diferentes cross-sections de los materiales. Es de especial interés, la presencia o no de fluorescencia, pues esta condicionará el tiempo que se necesite para adquirir un espectro de buena calidad. A su vez, ciertas muestras presentan una señal Raman muy débil y para lograr que mejore se deben buscar puntos minerales muy cristalinos. También ayuda en estos casos, poder analizar puntos que han estado protegidos de la erosión superficial, para lo cual puede ser útil fracturar las muestras y adquirir espectros de su parte interior.

3. TRABAJO COMPUTACIONAL: DETECCIÓN AUTOMÁTICA DE PICOS RAMAN E IDENTIFICACIÓN DE MATERIALES BASADA EN LA MULTIPLICACIÓN DE ESPECTROS.

El objetivo general de esta parte del trabajo es obtener un método de clasificación espectral que permita, de forma automática, realizar la clasificación e identificación de los materiales detectados en un espectro utilizando como única entrada el espectro en bruto de una determinada muestra problema. El trabajo se realiza de forma secuencial siguiendo los pasos lógicos de ejecución de la rutina: 1- procesamiento espectral para mejorar la calidad espectral; 2- detección automática de picos y 3- método de clasificación espectral.

Por tanto, en primer lugar se presenta un estudio del efecto que tiene sobre los espectros realizar acumulaciones mediante la multiplicación en lugar del habitual promediado aritmético. Para ello, se evaluarán en un primer lugar las ventajas reales que ofrece este método, para después implementar la rutina óptima de procesamiento espectral por multiplicación.

Tras hallar el procedimiento óptimo para la acumulación en base a multiplicación, se describe un algoritmo diseñado para realizar la detección automática de picos Raman basado en esta acumulación. El problema principal a resolver es establecer un umbral que marque el límite entre la señal Raman y el ruido espectral. Ha de tenerse en cuenta que los espectros con los que se trabaja tienen unas características muy específicas que están asociadas al espectrómetro con el cual han sido adquiridos. Es por ello por lo que se han establecido criterios para evitar la sobre detección de picos en base a características propias del espectro bajo estudio en lugar de en parámetros externos.

Finalmente, el método de detección automática se integra con un algoritmo que permite identificar un material gracias a su señal Raman, mediante comparación con información contenida en base de datos. Dicho trabajo de integración incluye la optimización y parametrización de dicho algoritmo para adaptarse de forma lo más universal posible a los diferentes escenarios de identificación que se han planteado. Este método de identificación ha sido probado extremo a extremo en un proceso que emula el tratamiento automático de los datos por parte de las herramientas de análisis que será realizado por RLS.

3.1. MATERIALES

Los materiales empleados en esta parte del estudio son:

- Espectros del modelo de sustitución de vuelo del instrumento RLS (FS, Flight Spare).
- Bases de datos de información espectral
 - Analytical Database of Martian Minerals (ADAMM).
 - Base de datos de estándares.
- Software de procesamiento espectral y desarrollo de algoritmia
 - (IDAT)/SpectPro
 - MATLAB
- Desarrollo existente del algoritmo de correlación espectral: identificación de minerales en base a picos Raman.

3.1.1. Espectros del modelo de sustitución de vuelo del instrumento RLS (FS, Flight Spare)

El modelo de sustitución de vuelo del instrumento RLS posee características prácticamente idénticas a las del modelo de a bordo. Para llevar a cabo el estudio presentado en este trabajo, orientado a dar soporte a la misión ExoMars, es de interés emplear espectros lo más parecidos posible a aquellos que llegarán de Marte. Adicionalmente, las muestras adquiridas por el RLS FS cuentan con más de un espectro por punto lo cual permite efectuar la multiplicación de espectros sobre un punto de la muestra, requisito indispensable para poder realizar esta parte del trabajo.

Las muestras empleadas para realizar el estudio sobre espectros de RLS FS son las siguientes:

- Hidróxido de Aluminio ($Al(OH)_3$)
- Azurita
- Calcita
- Hidróxido de Calcio ($Ca(OH)_2$)
- Tetracloruro de Carbono (CCl_4)
- Ciclohexano
- Diamante
- Indeno
- Vermiculita

El instrumento FS se encuentra ubicado en la sede del Instituto Nacional de Técnica Aeroespacial (INTA), en Torrejón de Ardoz, Madrid. Los espectros no han sido adquiridos por el autor del trabajo, sino que se utilizaron los espectros obtenidos durante una campaña de evaluación de prestaciones científicas.

3.1.2. Bases de datos de información espectral

Analytical database of Martian Minerals (ADAMM) La base de datos ADAMM ha sido ampliada en el desarrollo de este trabajo. Los objetivos de esta base de datos y la contribución realizada en ella pueden consultarse en la sección 2.3.

Base de datos de estándares La base de datos de estándares es una base de datos adicional a ADAMM desarrollada por el grupo de investigación ERICA.

Cuenta con espectros Raman tomados con distintas longitudes de excitación ($\lambda_1 = 532nm$ $\lambda_2 = 633nm$ $\lambda_3 = 1064nm$), espectros infrarrojos y de rayos-X. Las muestras que contiene son 53 minerales y moléculas de interés para el estudio geológico.

- | | | | |
|--------------|--------------|-----------------|---------------|
| ▪ Actinolita | ▪ Celestina | ▪ Grosularia | ▪ Piroxeno |
| ▪ Alunita | ▪ Copiapita | ▪ Halita | ▪ Rodocrosita |
| ▪ Anatasa | ▪ Crisotilo | ▪ Hematita | ▪ Rozenita |
| ▪ Anglesita | ▪ Crocoita | ▪ Hidrocerusita | ▪ Rutilo |
| ▪ Anhidrita | ▪ Cromita | ▪ Jarosita | ▪ Serpentina |
| ▪ Antigorita | ▪ Cuarzo | ▪ Lepidolita | ▪ Siderita |
| ▪ Apatito | ▪ Dolomita | ▪ Magnesita | ▪ Silicio |
| ▪ Aragonito | ▪ Epidota | ▪ Magnetita | ▪ Sodalita |
| ▪ Azufre | ▪ Esfalerita | ▪ Manganeso | ▪ Talco |
| ▪ Barita | ▪ Fluorita | ▪ Sulfato de Mg | ▪ Vermiculita |
| ▪ Berilo | ▪ Forsterita | ▪ Sulfato de Na | ▪ Yeso |
| ▪ Calcita | ▪ Galena | ▪ Ortoclasa | |
| ▪ Caolinita | ▪ Goethita | ▪ Pirita | |
| ▪ Carbón | ▪ Grafito | ▪ Pirolusita | |

Los espectros de ambas bases de datos han sido combinados y en ciertas ocasiones se han agrupado en función de su grupo mineral. La clasificación de minerales en grupos es especialmente útil en el caso de soluciones sólidas donde identificar el mineral concreto de la solución no es sencillo pues puede diferir de otro en pequeños porcentajes compositivos.

3.1.3. Software de procesamiento espectral y desarrollo de algoritmia

(IDAT)/SpectPro La herramienta de análisis espectral (IDAT)/SpectPro ha sido desarrollada por el grupo ERICA de la Universidad de Valladolid con el fin de recibir, decodificar, calibrar y verificar las telemetrías generadas por el instrumento RLS en Marte. Además, permite realizar el tratamiento y procesamiento espectral de los datos recibidos del instrumento, para lo que SpectPro cuenta con diversas funciones de tratamiento espectral como son el corte de espectros, la eliminación de línea de base, el ajuste de bandas, la medida de la relación señal-ruido, etc.

Asimismo, SpectPro está diseñado para trabajar con datos recibidos del instrumento RLS, de forma que ha sido utilizado para trabajar con los espectros adquiridos por el instrumento RLS-FS, ya que es necesario convertir los archivos contenidos en un fichero de sesión almacenado en un fichero .rls. La sesión .rls contiene, además de la información las telemetrías existentes en el instrumento durante el proceso de adquisición de espectros, los espectros (imágenes bidimensionales) en píxel adquiridos por la CCD del instrumento. Haciendo uso de las herramientas de procesamiento incluidas en SpectPro es posible calibrar los espectros y transformarlos en espectros unidimensionales fácilmente interpretables.

MATLAB Entorno de programación y cálculo numérico creado por Mathworks. Permite una cómoda manipulación de datos mediante el trabajo con matrices. Su nombre es la abreviatura de MATrix LABoratory, y se ha utilizado para el tratamiento de espectros y el desarrollo de algoritmia. Su principal ventaja frente a SpectPro en el tratamiento espectral es que permite manejar más cómodamente una cantidad elevada de espectros, lo cual es un requerimiento necesario para este trabajo. A su vez, ha sido el entorno donde se ha desarrollado el algoritmo de identificación automática de picos Raman y donde se ha modificado el código de detección automática que se presenta a continuación.

3.1.4. Desarrollo existente del algoritmo de correlación espectral: identificación de minerales en base a picos Raman

Durante el curso 2019-2020, Mohamed Khadmaoui Bichouna creó para su Trabajo de Fin de Grado: *Desarrollo de algoritmia para la identificación automática de espectros Raman de la misión ExoMars* (16) un algoritmo capaz de comparar la información espectral (posición e intensidad de picos Raman) obtenida manualmente de los espectros a identificar, con aquella contenida en espectros de referencia pertenecientes a una base de datos. De esta forma, se realiza una clasificación del espectro problema obteniendo una serie de posibles muestras coincidentes, ordenados según un indicador de coincidencia parametrizado por cinco variables que permiten modular el comportamiento del algoritmo:

- Tolerancia en intensidad
- Tolerancia en posición
- Peso de la intensidad
- Peso de la posición
- Peso del número de bandas

Buscando optimizar el método, M. Khadmaoui creó una serie de rutinas que probaban distintas configuraciones de estos parámetros por fuerza bruta hasta encontrar la combinación que proporcionaba el mayor número de aciertos y el menor de falsos positivos para un determinado set de muestras problema. La utilidad de esta optimización se ha mantenido y la modificación realizada sobre ella atiende de forma concienzuda y estructurada a mejoras que se detallarán más adelante.

El algoritmo final de M. Khadmaoui proporcionó un 100% de aciertos y solo un 30% de falsos positivos para muestras puras y un 75% de aciertos y 5% de falsos positivos para muestras mezcla. Sin embargo, los resultados se realizaron sobre un conjunto relativamente pequeño de muestras problema y de la base de datos, por lo que estas cifras han de ser manejadas con una cierta precaución.

El código creado por M. Khadmaoui para su Trabajo de Fin de Grado ha sido empleado y modificado durante este estudio, como se recoge en la sección 3.4. Dicho algoritmo consiste en una rutina de clasificación que efectúa la detección automática de minerales mediante comparación con información contenida en base de datos. Para ello sigue los siguientes pasos:

1º Lectura de datos del espectro problema y la base de datos. El algoritmo de clasificación comienza abriendo los archivos .bds (un tipo de fichero XML, ver Figura 32.) que contienen la información de posición e intensidad ($\langle bandPosition \rangle$ y $\langle bandIntensity \rangle$) de los picos de los espectros.

2º Selección de un subconjunto de espectros de la base de datos considerados posibles candidatos. Esta selección se realiza en base a la identificación y comparación del pico principal del problema contra los espectros de la base de datos. Este paso es necesario para aligerar por un lado la carga computacional del algoritmo, pero sobre todo para evitar la aparición de falsos positivos: al reducir los candidatos de coincidencia a un subconjunto de espectros, las probabilidades de encontrar múltiples coincidencias de picos se reducen. Esta reducción del conjunto de espectros candidatos tiene su razón de ser en el hecho de clasificar los espectros en base a su pico principal, ya que no tiene sentido calcular un porcentaje de coincidencia con aquellas fases minerales cuya vibración principal no coincida con la del espectro problema. Para ello, el algoritmo inicial de M. Khadmaoui, comienza normalizando el espectro problema y buscando el pico principal del mismo, es decir, aquel pico cuya intensidad valdrá 1. Posteriormente, se busca de entre todos los picos principales de los materiales de la base de datos aquellos que coincidan con el pico principal del espectro problema con un rango de tolerancia de 9cm⁻¹ (parámetro de tolerancia de posición que M. Khadmaoui obtuvo en su estudio del método).

```

<?xml version="1.0" encoding="UTF-8"?>
<ListFittingBands>
  <Band>
    <bandName>Band 1</bandName>
    <bandPosition>129.22</bandPosition>
    <bandIntensity>0.55</bandIntensity>
    <bandFWHM>13.60</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 2</bandName>
    <bandPosition>245.89</bandPosition>
    <bandIntensity>0.23</bandIntensity>
    <bandFWHM>14.50</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 3</bandName>
    <bandPosition>268.45</bandPosition>
    <bandIntensity>0.37</bandIntensity>
    <bandFWHM>19.49</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 4</bandName>
    <bandPosition>335.29</bandPosition>
    <bandIntensity>0.34</bandIntensity>
    <bandFWHM>15.09</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 6</bandName>
    <bandPosition>427.72</bandPosition>
    <bandIntensity>0.34</bandIntensity>
    <bandFWHM>23.16</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 7</bandName>
    <bandPosition>465.64</bandPosition>
    <bandIntensity>0.54</bandIntensity>
    <bandFWHM>24.19</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 8</bandName>
    <bandPosition>3622.10</bandPosition>
    <bandIntensity>1.00</bandIntensity>
    <bandFWHM>7.52</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 9</bandName>
    <bandPosition>3692.06</bandPosition>
    <bandIntensity>0.67</bandIntensity>
    <bandFWHM>21.11</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 10</bandName>
    <bandPosition>199.07</bandPosition>
    <bandIntensity>0.19</bandIntensity>
    <bandFWHM>25.39</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 11</bandName>
    <bandPosition>512.83</bandPosition>
    <bandIntensity>0.14</bandIntensity>
    <bandFWHM>18.39</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 12</bandName>
    <bandPosition>704.88</bandPosition>
    <bandIntensity>0.13</bandIntensity>
    <bandFWHM>21.39</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 13</bandName>
    <bandPosition>749.52</bandPosition>
    <bandIntensity>0.17</bandIntensity>
    <bandFWHM>32.39</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 14</bandName>
    <bandPosition>787.60</bandPosition>
    <bandIntensity>0.14</bandIntensity>
    <bandFWHM>14.39</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 15</bandName>
    <bandPosition>914.95</bandPosition>
    <bandIntensity>0.12</bandIntensity>
    <bandFWHM>25.39</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 16</bandName>
    <bandPosition>939.72</bandPosition>
    <bandIntensity>0.08</bandIntensity>
    <bandFWHM>10.39</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 17</bandName>
    <bandPosition>3652.88</bandPosition>
    <bandIntensity>0.20</bandIntensity>
    <bandFWHM>18.76</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 18</bandName>
    <bandPosition>3669.58</bandPosition>
    <bandIntensity>0.15</bandIntensity>
    <bandFWHM>8.76</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
</ListFittingBands>

```

Figura 32: Ejemplo de información del ajuste de bandas de la caolinita. La muestra empleada en este caso es una caolinita procedente de Orleans.

3º Cálculo del factor de coincidencia. Aquellos materiales que cumplieran la condición de coincidencia de picos principales pasaban a ser considerados en una lista que luego permitía explorar cual de ellos coincidía mejor con el espectro problema mediante el cálculo del factor de coincidencia.

La función de coincidencia analiza cada uno de los candidatos propuestos previamente comparando todos sus picos Raman con los presentes en el espectro problema. Una vez establecida la semejanza en posición de los picos (usando de nuevo el parámetro *tolerancia_posición*) se verifica que las intensidades de ambos sean similares. El valor de la tolerancia en posición es 0,4.

Para ordenar la coincidencia de los diversos candidatos se crean dos vectores: intensidad y posición en los que se guarda el valor absoluto de la diferencia entre posiciones e intensidades de los distintos picos emparejados del espectro problema y la muestra correspondiente de la base de datos.

Posteriormente, se guarda el número de picos emparejados y se multiplica el valor medio del vector intensidad y el valor medio del vector posición por el peso dado a la intensidad (0,2) y el peso dado a la posición (0,2), respectivamente.

Se normaliza el número de picos emparejados al máximo hallado de entre todas las muestras y se asigna el valor cero a este material. El resto tendrán la unidad menos su valor normalizado. Se multiplica este parámetro modificado del número de picos por el peso dado al número de bandas (0,6). Finalmente, calcula el parámetro de coincidencia como se muestra en la ecuación 22.

$$\text{coincidencia} = (1 - (P_i * p_i + P_p * p_p + P_b)) * 100 \quad (22)$$

donde P_i es el promedio del vector intensidad, p_i el peso dado a la intensidad, P_p el promedio del vector posición, p_p el peso dado a la posición y P_b el parámetro modificado de bandas.

De este modo la muestra de la base de datos que tenga mayor semejanza con el espectro problema tendrá un valor de coincidencia mayor al resto por tener los menores valores de promedio de vector intensidad, promedio de vector posición y parámetro modificado de bandas.

4º Clasificación. En base al parámetro de coincidencia calculado, se clasifican los candidatos en orden decreciente de probabilidad de coincidencia, obteniendo un candidato óptimo para el espectro problema en cuestión. EL algoritmo muestra asimismo el valor calculado de coincidencia.

5º Actualización de la información de bandas. Una vez detectado un material, se procede a eliminar la contribución del espectro detectado del espectro problema, restando la contribución en intensidad de todas las bandas coincidentes del espectro problema. De esta forma es posible realizar una nueva iteración del algoritmo con las bandas restantes del espectro, dotando al proceso de la capacidad de detectar la presencia de uno o varios materiales presentes en el espectro analizado.

6º Búsqueda de nuevos materiales. Para realizar una nueva iteración en busca de otros posibles materiales, se toma el espectro al que se le ha restando la contribución del elemento detectado, se renormaliza (intensidad del pico mayor igual a 1), y se ejecuta una nueva iteración del algoritmo. Este proceso recursivo se detiene cuando no se encuentran nuevas coincidencias, es decir, cuando el espectro problema tras el paso 5º es igual en dos iteraciones consecutivas.

3.2. DEFINICIÓN DEL MÉTODO DE ACUMULACIÓN POR MULTIPLICACIÓN

3.2.1. Estudio preliminar

Al comenzar a explorar esta vía de tratamiento y acumulación espectral se vio que la multiplicación de los espectros promocionaba los picos de señal Raman varios órdenes de magnitud por encima del ruido del espectro. Esta observación inicial se llevó a cabo usando SpectPro y una muestra de ciclohexano adquirida por el RLS FS. En base a esto se decidió realizar un análisis más profundo utilizando todas las muestras adquiridas por el RLS FS (consultar sección 3.1.1).

Una de las primeras observaciones realizadas sobre el ciclohexano (véase Figura 33.) fue que al multiplicar los espectros solo se veían los picos principales por tener una intensidad varios ordenes de magnitud superior al resto de picos del espectro. Sin embargo, al reducir la escala de observación se veía que sucedía algo similar con los picos secundarios, que quedaban aún varios ordenes de magnitud por encima del ruido.

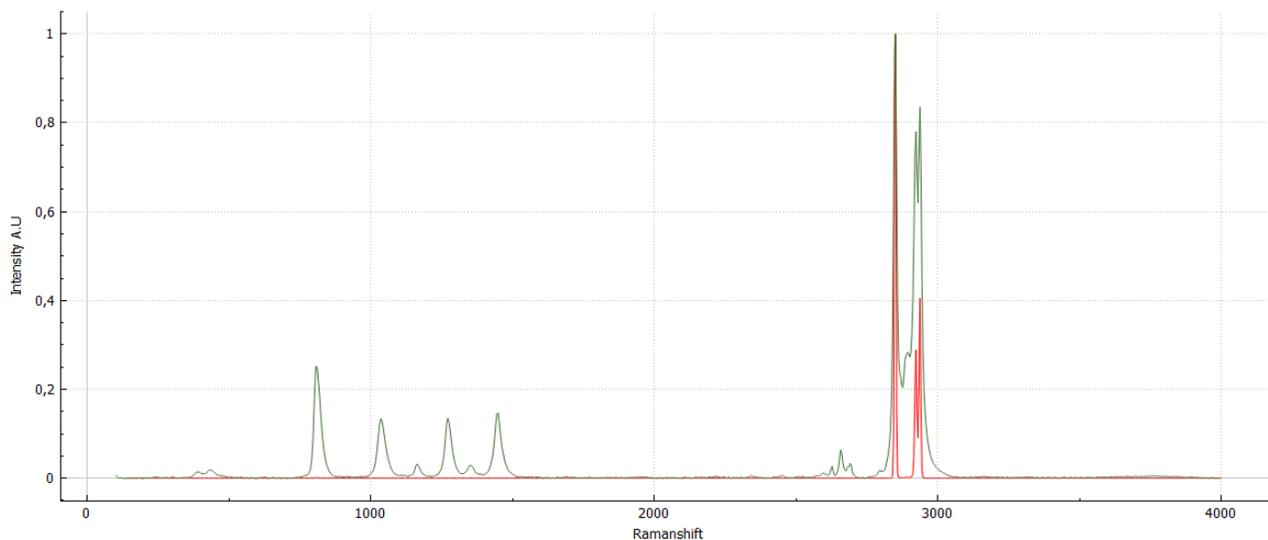
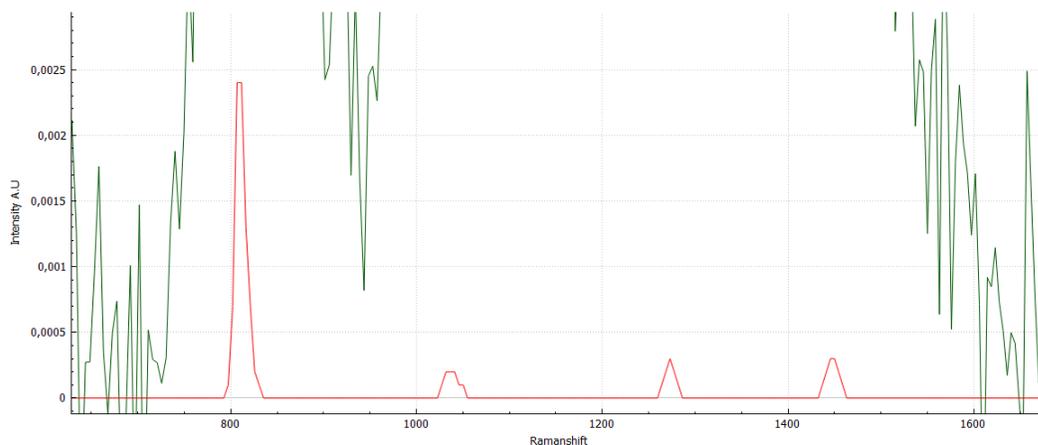
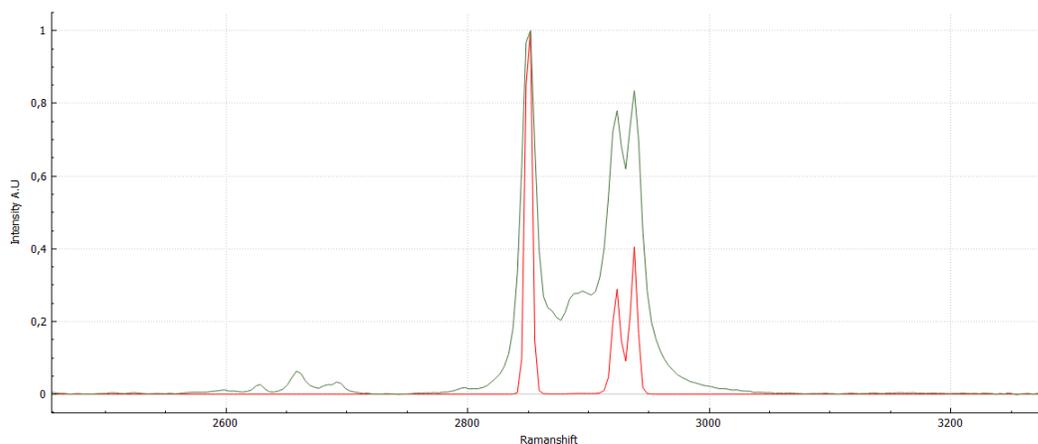


Figura 33: Espectro de ciclohexano. En rojo se muestra la acumulación por multiplicación y en verde el típico promedio aritmético. Ambos espectros han sido normalizados para su visualización.



(a) Ampliación en la zona baja del espectro del ciclohexano. En rojo se muestra la acumulación por multiplicación y en verde el típico promedio aritmético.



(b) Ampliación en la zona alta del espectro del ciclohexano. En rojo se muestra la acumulación por multiplicación y en verde el típico promedio aritmético.

Figura 34: Comparativa del espectro de ciclohexano. Acumulación por multiplicación frente a promedio aritmético.

En la Figura 34, con los espectros obtenidos con ciclohexano, se aprecia cómo se produce una importante disminución relativa de la intensidad de los picos de la parte baja del espectro multiplicado a causa de la diferencia de órdenes de magnitud que produce la multiplicación. Esta disminución es tal que dichos picos ni siquiera son perceptibles en la imagen, pero son perfectamente detectables si se amplía lo suficiente (ver Figura 34a.). Al ampliar la imagen también puede observarse que no aparece ruido en el espectro multiplicación, mientras que sí lo hace en el espectro suma (ver Figura 34b.).

Es muy interesante también destacar como los picos de la zona alta (2853cm^{-1} , 2924cm^{-1} y 2938cm^{-1}) se estrechan e incluso separan, en un proceso conocido como autoresolución facilitado por la multiplicación espectral (véase Figura 34b.).

Si bien aparentemente la multiplicación de espectros resulta en incrementos notorios de la SNR del espectro, es cierto también que no es fácil visualizar toda la información espectral en una imagen del espectro. Por tanto, para poder visualizar los picos secundarios (con intensidades órdenes de magnitud menores) se creó un sencillo programa que permitía observar con detalle las diversas zonas del espectro, para así alejarnos de los picos principales cuya promoción impide la visualización del resto.

Repitiendo el tratamiento con otras muestras, se observaron comportamientos dispares. Concretamente, con los Hidróxidos de Al (Figura 35.) y Ca (Figura 36.), puede apreciarse cómo la línea de base se promociona notablemente de tal modo que enmascara la información espectral. Esta característica no se observaba con el espectro de ciclohexano mostrado como ejemplo por no contar este con línea de base. Por tanto, al presentarse este comportamiento de crecimiento descontrolado de la línea de base en varios espectros, se pone de manifiesto que es un problema necesario de solucionar para que este método ofrezca alguna ventaja.

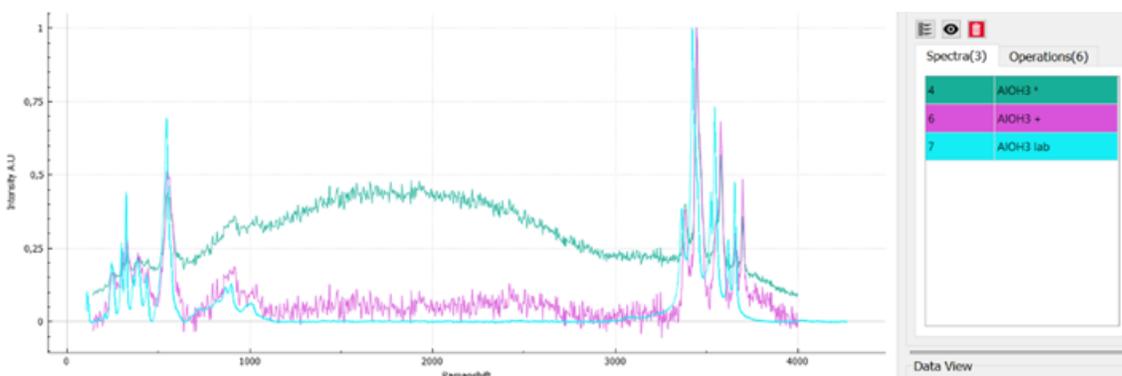


Figura 35: Espectros de hidróxido de aluminio suma, en rosa, multiplicación en verde y referencia de laboratorio en azul. Los espectros han sido normalizados para poder llevar a cabo la comparación de forma cómoda.

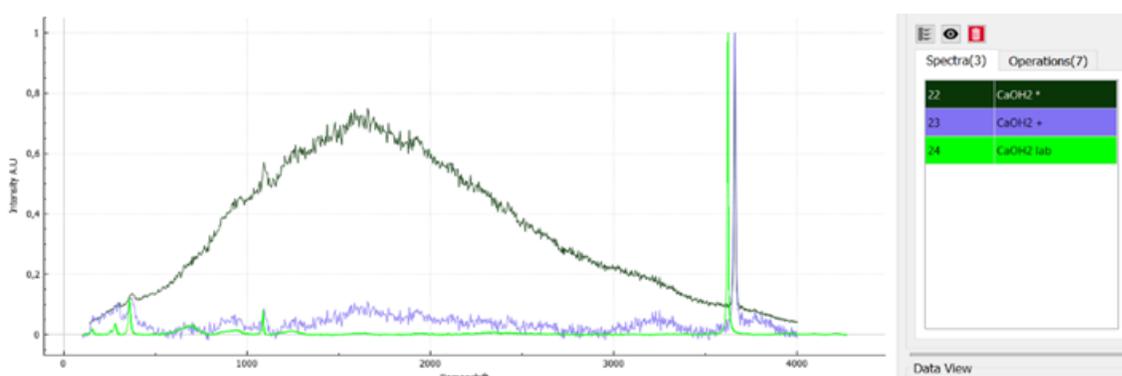


Figura 36: Espectros de hidróxido de calcio suma, en azul, multiplicación en verde oscuro y referencia de laboratorio en verde brillante. Los espectros han sido normalizados.

En todo caso, se realizó un estudio de la relación señal ruido de los espectros, recogido en la tabla de la Tabla 7., donde se observa que la mejora es muy notoria en aquellos espectros en los que el crecimiento de la línea de base no es significativo y presentan picos claros como son el ciclohexano, el diamante y el silicio. Por tanto, se concluye que para que la multiplicación de espectros sea efectiva, será necesario realizar una correcta eliminación de la línea de base para mejorar el método.

En este punto del estudio se probaron diversos métodos y aproximaciones para proceder a la eliminación de línea de base. Cuando se trabaja con un espectro suma, lo habitual es eliminar la línea de base del espectro final, ya que se facilita el promediado de espectros similares adquiridos en iguales condiciones. Sin embargo, si se elimina la línea de base de los espectros individuales y luego se realiza el promedio aritmético, se acumulan los errores producidos por el proceso de eliminación de la línea de base en comparación con la eliminación de la línea de base del espectro promedio. En el caso de la multiplicación espectral, la opción de eliminar la línea de base después de la multiplicación no aporta beneficios notables. Pero sí que se realizaron múltiples pruebas y combinaciones de procesos (con muchos caminos infructuosos) para combinar la eliminación de la línea de base, la corrección en intensidad y la función aparato mediante la división entre un espectro amorfo de calibración.

Muestra	Región de ruido	Acumulación por multiplicación		Promedio aritmético	
		Rango de pico	SNR	Rango de pico	SNR
Ciclohexano	1700-1900	777.96-845.92	43045.1	758.53-876.50	379.334
Ciclohexano	1700-1900	1020.19-1051.87	3001.17	989.90-1092.78	196.143
Ciclohexano	1700-1900	2821.28-2877.74	1.80E+07	2804.07-2878.92	1312.68
$Al(OH)_3$	1500-1700	473.28-638.35	11.3527	491.73-639.57	15.4015
$Al(OH)_3$	1500-1700	3412.05-3507.43	35.5617	3407.95-3516.37	29.1447
Indeno	1910-2110	1597.41-1623.95	0.15136	1576.37-1637.59	3.31646
Indeno	1910-2110	3034.71-3087.22	0.271619	2972.05-3090.99	3.17583
Silicio	1100-1300	496.50-565.46	22.4605	496.32-573.49	15.3056
Azurita	1750-1950	2362.44-2494.47	772	2374.83-2492.53	26.8204
Calcita	1500-1700	1062.17-1131.00	547.685	1055.84-1145.12	221.255
$Ca(OH)_2$	2800-3000	3621.01-3692.07	56.7811	3620.61-3690.34	82.0485
CCl_4	1000-1200	198-254.95	86286.4	179.27-274.71	405.385
CCl_4	1000-1200	293.31-359.70	137578	278.31-391.76	455.247
CCl_4	1000-1200	432.25-512.58	1.26E+06	407.14-550.85	798.229
Diamante	1680-1880	1304.70-1377.86	3246.12	1285.89-1389.27	513.348
Talco	1200-1400	643.65-738.08	3.40945	640.38-749.14	10.2625
Talco	1200-1400	3707.50-3740.85	2.48643	3705.64-3734.50	8.94574

Tabla 7: Estudio de la relación señal ruido. Comparación acumulación por multiplicación y promedio aritmético previo tratamiento espectral.

3.2.2. Método optimizado

De este modo, se consigue determinar lo que se ha considerado el método óptimo de acumulación basado en multiplicación de espectros. El método óptimo de tratamiento espectral consiste en que los espectros han de ser procesados previamente a realizar la multiplicación mediante la división por el espectro amorfo para eliminar la función aparato, y después se elimina la línea de base. Así se logra que el ruido asociado a la función aparato del instrumento y la línea de base no contribuyan ni se promocionen durante la creación del espectro multiplicación. De esta forma en el espectro final solo se observa la información Raman y el ruido residual del espectro que no ha podido ser eliminado. Como se verá a continuación, dicho ruido no tiene una contribución notable hasta un cierto límite de ordenes de magnitud.

El algoritmo optimizado queda por tanto definido de la siguiente manera:

- 1º Cortar los espectros al rango espectral de interés.
- 2º Dividir los espectros individuales entre el borato de zinc normalizado.
- 3º Eliminar la línea de base usando una curva muy noble mediante el método automático que proporciona SpectPro. Parámetros $\delta_1 = 20$, $\delta_2 = 30$, $BandLimit = 1200$ y $FilterIndex = 5$.
- 4º Multiplicar todos los espectros del mismo punto.

Para ilustrar el funcionamiento del algoritmo, se muestra a continuación el proceso llevado a cabo para el hidróxido de calcio ($Ca(OH)_2$).

En primer lugar, se cortan los espectros individuales de acuerdo con el rango en el que se pretende trabajar (ver Figura 37.). En este caso se busca aprovechar al máximo el rango espectral, por lo que se corta entre $140cm^{-1}$ y $4000cm^{-1}$. El corte en la zona baja se efectúa en un punto en cierto modo alto debido a que se quiere eliminar por completo la contribución residual del pico de dispersión Rayleigh antes de eliminar la línea de base.

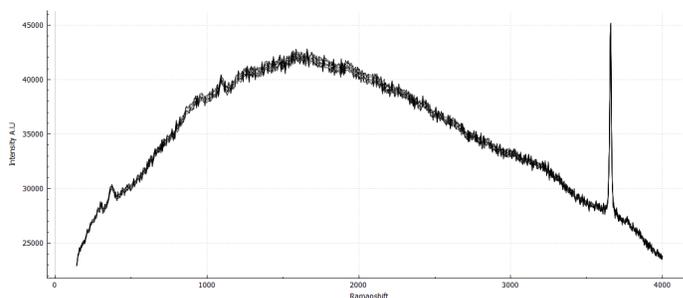


Figura 37: Espectros individuales de $Ca(OH)_2$ cortados en el rango de trabajo $140-4000cm^{-1}$.

En segundo lugar, se dividen los espectros recortados entre el espectro de borato de zinc normalizado, eliminando de forma efectiva el ruido y rizado asociados a la función aparato (ver Figura 38.).

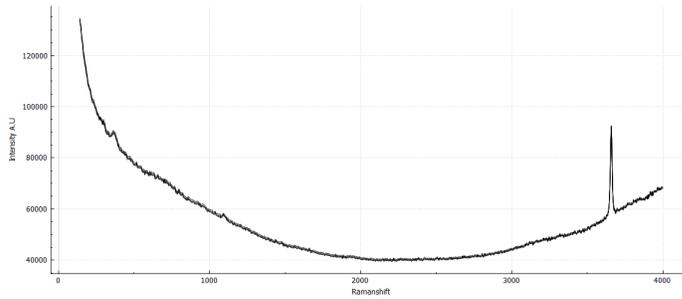


Figura 38: Espectros de hidróxido de calcio $Ca(OH)_2$ divididos entre el espectro de borato de zinc amorfo.

Se ve, al dividir los espectros problema entre el borato de zinc amorfo, que la curvatura de la línea de base cambia de signo y que los picos Raman aparecen con mayor claridad en relación al ruido espectral. El cambio de la curvatura se debe a la división por intensidades muy pequeñas en los límites del espectro. La mejor visualización de los picos Raman está causada por la eliminación de gran parte del ruido asociado a la función aparato.

Por último se elimina la línea de base de los espectros individuales (Figura 39.) y se multiplican, obteniendo los resultados de la Figura 40.. Como puede observarse, en la imagen solo se aprecia el pico principal situado en 3658cm^{-1} . Sin embargo, si ampliamos la imagen en la zona baja del espectro podemos observar el pico secundario del espectro (ver Figura 41.).

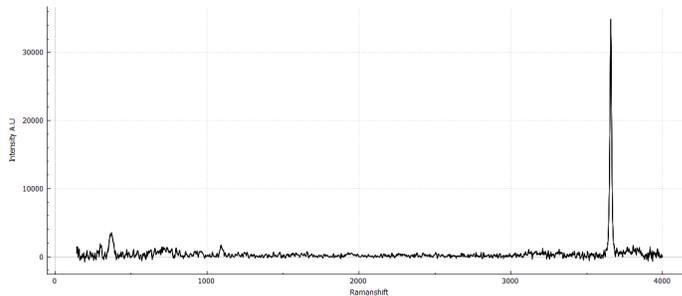


Figura 39: Espectros de hidróxido de calcio $Ca(OH)_2$ con la línea de base eliminada.

Finalmente, para cuantificar de forma efectiva la mejoría obtenida con este sistema optimizado de acumulación por multiplicación, se repite el estudio de la relación señal-ruido para todos los materiales, mostrando una clara mejoría en términos de SNR para todos los espectros en comparación con los valores obtenidos con el promedio aritmético (véase Tabla 8.)

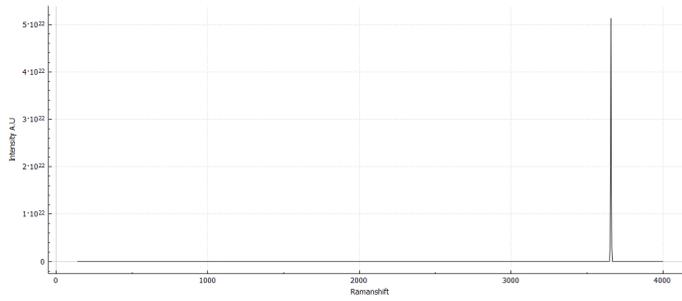


Figura 40: Espectro multiplicación del hidróxido de calcio $Ca(OH)_2$ creado a partir de espectros individuales tratados previamente.

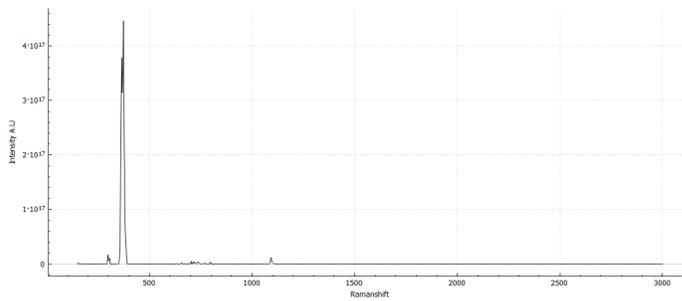


Figura 41: Espectro multiplicación de hidróxido de calcio $Ca(OH)_2$ en el rango $150 - 3000\text{cm}^{-1}$.

Muestra	Región de ruido	Acumulación por multiplicación		Promedio aritmético	
		Rango de pico	SNR	Rango de pico	SNR
Ciclohexano	1700-1900	775.94-850.03	5.70E+12	758.53-876.50	379.334
Ciclohexano	1700-1900	2820.31-2878.44	6.62E+14	2804.07-2878.92	1312.68
$Al(OH)_3$	1500-1700	473.79-639.27	5.82E+06	491.73-639.57	15.4015
$Al(OH)_3$	1500-1700	3410.55-3506.98	4.00E+07	3407.95-3516.37	29.1447
Indeno	1910-2110	1598.03-1624.32	7009.65	1576.37-1637.59	3.31646
Indeno	1910-2110	3035.75-3089.12	5360.53	2972.05-3090.99	3.17583
Silicio	1100-1300	497.27-565.46	781771	496.32-573.49	15.3056
Calcita	1500-1700	1071.04-1120.60	1.15E+11	1055.84-1145.12	221.255
$Ca(OH)_2$	2800-3000	3619.96-3692.37	3.93E+09	3620.61-3690.34	82.0485
CCl_4	1000-1200	199.95-247.63	4.52E+13	179.27-274.71	405.385
CCl_4	1000-1200	300.68-344.34	3.00E+13	278.31-391.76	455.247
CCl_4	1000-1200	438.04-503.35	1.81E+14	407.14-550.85	798.229
Diamante	1680-1880	1319.66-1355.75	2.36E+13	1285.89-1389.27	513.348
Talco	1200-1400	662.12-738.08	110402	640.38-749.14	10.2625
Talco	1200-1400	3706.85-3735.67	210722	3705.64-3734.50	8.94574

Tabla 8: Estudio de la relación señal ruido. Comparación acumulación por multiplicación y promedio aritmético tras llevar a cabo el tratamiento espectral.

3.3. IMPLEMENTACIÓN DEL ALGORITMO PARA LA DETECCIÓN AUTOMÁTICA DE PICOS

Este tipo de espectro obtenido por multiplicación proporciona un espectro que difícilmente puede ser visualizado adecuadamente para poder realizar un procesado por parte un operador, debido a la complicada representación de un rango tan grande de órdenes de magnitud. Por otro lado, se ha observado que el espectro promedio geométrico (la raíz del espectro multiplicación) no proporciona un resultado apreciablemente mejor que el obtenido con el promedio aritmético.

Por tanto, parece claro que la ventaja de este método estaría en utilizar el espectro multiplicado. Aunque de difícil representación para un operador humano, el hecho de que la SNR sea tan grande, facilita que rutinas de detección automática de picos puedan detectar dichos picos con mayor facilidad que en un espectro ruidoso. Por tanto, esto supone una situación muy ventajosa a la hora de implementar un algoritmo de detección automática de picos.

Un aspecto crítico a la hora de considerar el método válido es el garantizar que sea posible discernir suficientemente los picos del ruido espectral, de forma que el algoritmo disponga de un criterio fiable para detener su ejecución para que no se produzca la detección de ruido como señal Raman. Viendo como se ha mostrado hasta ahora, que era necesario reducir ordenes de magnitud en la búsqueda de picos secundarios en el espectro multiplicación, era difícil, en primera instancia, saber hasta cuando se debía reducir.

Para lograr establecer una condición segura se estudiaron en detalle los espectros con que se contaba del instrumento RLS-FS. En primer lugar, fue necesario aprender qué picos deberían aparecer en cada espectro, para saber qué es lo que debía detectar el algoritmo. Para ello, se llevó a cabo un trabajo bibliográfico (Figura 42.) que orientó las posteriores decisiones. Para contrastar dicho estudio se pueden consultar las referencias (17), (18), (19), (20), (21), (22), (23) y (24).

Picos Raman de los diversos materiales (cm-1)								
AlOH3	Azurita	Calcita	CaOH2	Ciclohexano	Diamante	Indeno	Vermiculita	Talco
250	154	155	359	364	1332	730	113	200
306	179	221	684	426		1018	227	297
322	281	282	712	801		1205	411	372
379	332	712	1086	1028		1552	643	682
434	332	221	3620	1158		1610	660	798
545	739	1025	684	1266		2892	678	3723
710	765	1067	712	1444		3054		
866	816	1086	1086	2664				
1019	1095	1749	3620	2853				
3364	1435	2148		2924				
3420		2215		2938				
3522								
3542								
3552								
3656								
3698								
3704								

Figura 42: Información de los picos Raman de los diversos materiales bajo estudio obtenida de la bibliografía.

Teniendo claro lo que debía detectarse se procedió a la búsqueda de una característica diferenciadora de los picos Raman que no poseyera el ruido espectral. Para ello, diversos caminos fueron explorados. En las siguientes secciones se exponen algunos de ellos y se explica con detenimiento el método final.

3.3.1. Requerimiento de promoción de los picos Raman

Al trabajar con el promedio geométrico se cuenta con varios espectros tomados sobre el mismo punto de la muestra. Serían las distintas acumulaciones del espectro, pero guardadas por separado. Dado que todos se corresponden con el mismo punto es de esperar que todos ellos contengan las señales Raman del material bajo estudio.

Un espectro individual contará con intensidades que mezclan el ruido y la señal espectral, de manera que al multiplicar varios espectros crezcan ambos. Los picos Raman de una misma muestra y un mismo punto tendrán un crecimiento homogéneo a medida que se produzca la multiplicación y recibiendo una contribución análoga de cada espectro. Sin embargo, es de esperar y así se ha visto en las muestras estudiadas, que el ruido no se proporcione del mismo modo por tener su posición e intensidad un cierto carácter aleatorio. El crecimiento del ruido será por tanto relativamente menor que el de los picos.

Esta forma de crecimiento del ruido impide establecer un criterio unánime para todo el espectro. Para solucionar esta cuestión se propone un criterio individualizado para cada punto del espectro que se base en cómo se ha comportado dicho punto frente al crecimiento por multiplicación. Es decir, se impondrá un criterio para cada punto del eje x del espectro (desplazamiento Raman) en el cual su intensidad asociada ha de llegar a la potencia n-ésima del valor del primer espectro del conjunto salvo un cierto margen de tolerancia.

Este método acabó fallando porque guardaba una fuerte dependencia con el espectro inicial usado para establecer el condicional. Tampoco fue posible establecer un margen de tolerancia que incluyera los picos sin dejar pasar puntos de ruido. Lo que sucedía es que solo ciertos puntos del pico pasaban la condición y éste se veía fraccionado (ver Figura 43.). Este comportamiento no guarda correspondencia con la forma de la emisión Raman que cuenta con una cierta anchura.

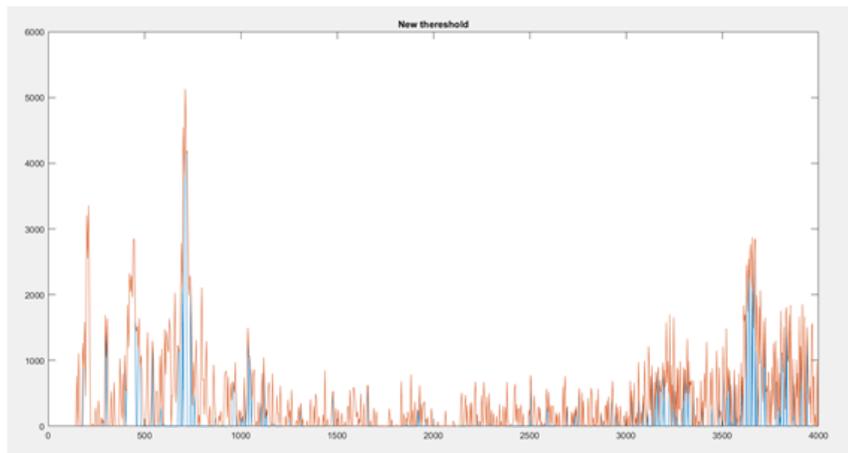


Figura 43: Fraccionamiento de los picos Raman a causa del establecimiento del umbral por crecimiento.

Por todas estas razones este método no se considera válido para la detección automática.

3.3.2. Método basado en los espectros estadísticos de las muestras de ensayo

Para analizar que diferencia a los picos Raman del ruido se decidió llevar a cabo un estudio estadístico de los espectros. Para ello, se creó un script de que calculaba diversos espectros estadísticos a partir de los espectros usados para la multiplicación. Estos espectros eran:

- Espectro mediana
- Espectro desviación estándar
- Espectro desviación estándar entre mediana
- Espectro de máximos
- Espectro de mínimos
- Espectro diferencia entre máximos y mínimos

A continuación, se muestran los resultados obtenidos para el caso del hidróxido de aluminio (ver Figura 44.).

Con estos resultados, se concluye que el estudio estadístico de los espectros llevado a cabo comparando unos con otros no da pistas útiles para la identificación de picos, y por tanto se descarta continuar por esta vía.

No obstante, como la presencia de los picos está caracterizada por su intensidad y anchura, queda claro que el condicional que se establezca para definir el final de las iteraciones del algoritmo ha de tener en cuenta la anchura e intensidad de los picos.

Adicionalmente, el espectro desviación estándar resulta muy interesante pues muestra como la variabilidad que sufren los picos es más acusada en los extremos del espectro. Sin embargo, su intensidad se mantiene más constante en la zona central. Este aspecto se venía notificando desde hace tiempo y ahora el espectro desviación estándar lo pone de manifiesto con mayor claridad.

A un tiempo que el script muestra los espectros estadísticos, también proporciona los valores de desviación estándar, mediana y desviación estándar entre mediana del espectro multiplicación (véase Tabla 9.).

Mediana	std	std/mediana
$8,23 \cdot 10^{11}$	$3,87 \cdot 10^{20}$	$4,70 \cdot 10^8$

Tabla 9: Valores mediana, std, std/mediana para el conjunto de espectros de hidróxido de aluminio.

Como puede observarse, el valor de la desviación estándar es un valor elevado, cercano en órdenes de magnitud a los picos que logran visualizarse en el espectro multiplicación. Es por ello por lo que este parámetro va a ser empleado en el establecimiento de un umbral para la intensidad, tal y como se verá más adelante.

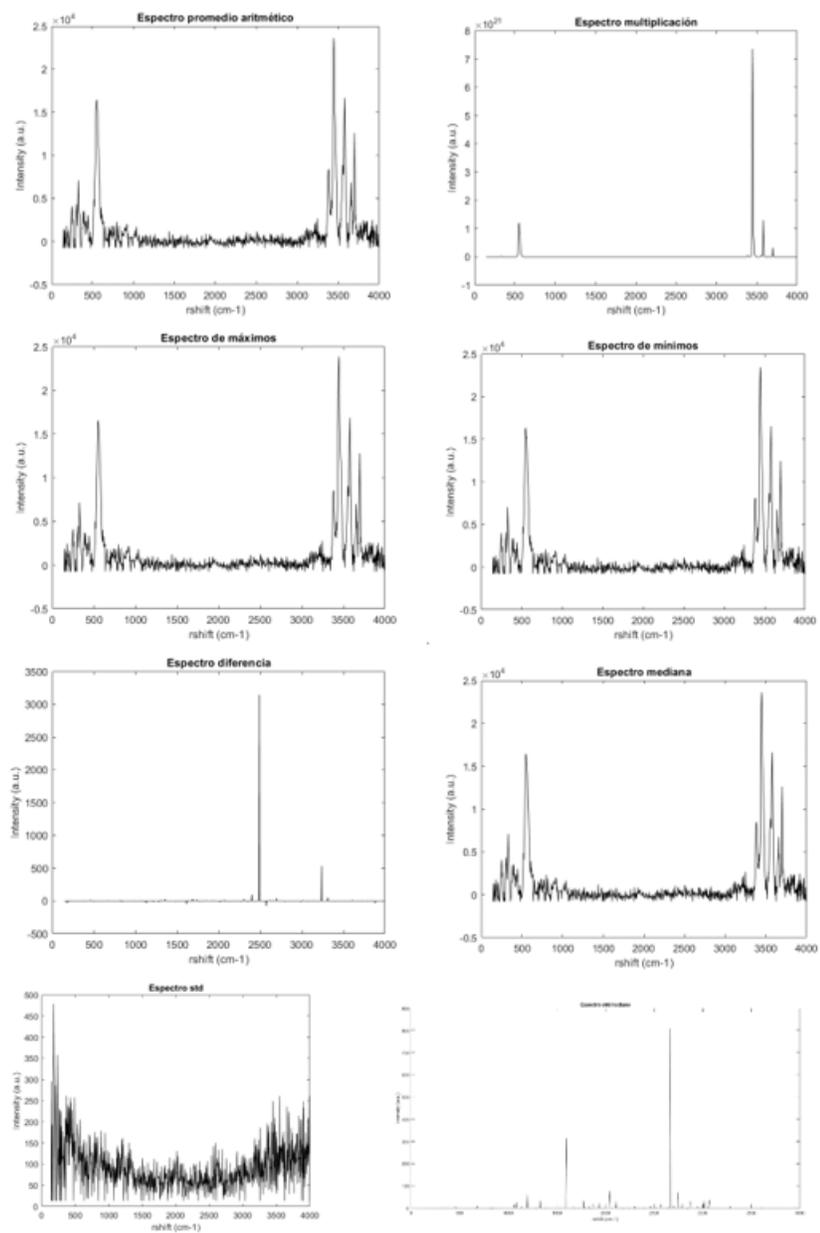


Figura 44: Conjunto de espectros estadísticos calculados en el estudio.

3.3.3. Sustitución por ceros, media del espectro o redimensión del vector.

Teniendo en cuenta las conclusiones obtenidas acerca de la idoneidad de utilizar los valores de intensidad y anchura de los picos como criterios para distinguir picos Raman de ruido, se decide tomar la intensidad como criterio para la detección de picos. Un criterio ampliamente extendido para considerar un pico suficientemente intenso establece que su intensidad ha de superar por tres veces el valor de la desviación estándar del espectro.

Con este método se ve que en una primera iteración se logran identificar uno o dos picos, una cantidad muy escasa. Para que la detección sea interesante, es necesario poder identificar más bandas. Por ello, se aplica el criterio recursivamente. Por tanto, sucesivas iteraciones recalcularán la desviación estándar, que deberá ignorar la intensidad de los picos que ya han sido detectados para poder avanzar. Es en este punto donde surgen tres caminos para el nuevo cálculo de la desviación estándar:

- Sustituir la intensidad de los puntos detectados por 0, de manera que no puedan ser detectados de nuevo.
- Sustituir la intensidad por el valor de la media del espectro, para que la sustitución de estos valores no altere significativamente el cálculo de la std.
- Crear un nuevo vector de menor dimensión que no contenga los picos ya detectados y continuar reduciéndolo iterativamente.

Tras probar todas las opciones, se observó que los tres métodos ofrecían resultados similares. Se detectaban muchos falsos positivos y algunos de los picos característicos de los espectros problema aparecían pasadas varias iteraciones. La cantidad de picos acertados detectados por los tres métodos era muy similar, si bien es cierto que el método de sustitución por la media proporcionaba más falsos positivos que los otros dos. Finalmente, se optó por continuar el estudio con el procedimiento de sustitución por ceros, ya que, ofreciendo resultados muy similares, resulta más sencillo computacionalmente.

3.3.4. Método final

Definición del método Con todo lo aprendido en los apartados precedentes, se decide implementar el método que sustituye los picos más intensos que tres veces la desviación estándar del espectro por ceros de forma iterativa. Ante la búsqueda de mejorar los resultados hallados, hay tres problemas que deben considerarse:

- El método detecta muchos falsos positivos en iteraciones anteriores a la detección correcta de alguno de los picos secundarios.
- La detección finaliza cuando no quedan picos con intensidad tres veces superior a la desviación estándar, con el consiguiente riesgo de sobredetectar picos.
- La detección de falsos positivos es acusadamente más elevada en la zona de grandes desplazamientos Raman del espectro, debido también a un mayor ruido y el efecto de la corrección de la línea de base.

El efecto de estos problemas puede verse en la Figura 45. donde se muestra el estado de la detección para el caso del ciclohexano.

Ciclohexano		
Método de sustitución por ceros		
Rshift	Intensidad	Iteración
301,8459	231,5510875	17
432,3594	1756,480432	10
629,1422	199,7927141	18
806,9736	16877,42418	3
1036,5057	7165,622781	5
1164,7109	1419,624209	10
1272,9396	5994,033609	6
1450,0012	5956,605818	6
2218,5556	189,0636708	18
2343,6705	219,1310102	17
2451,3759	211,5816709	17
2658,2744	2617,106006	8
2852,0042	44931,23871	1
3158,5906	188,2585482	18
3168,9185	207,5232087	17
3672,144	167,358887	19
3732,398	191,7409174	18
3767,0246	333,7143489	15
3788,9623	224,8990781	17
3804,586	242,2838339	17
3820,1714	187,8852501	18
3838,8236	233,8819275	17
3848,1292	263,4812832	16

Figura 45: Muestra de la sobredetección que tiene lugar cuando el método no ha sido mejorado. El ejemplo corresponde al ciclohexano. En verde, los picos Raman de este material.

Para solventar los problemas presentados se plantean varias modificaciones y criterios de mejora adicionales. Los criterios se resumen a continuación.

Criterio 0 (de parada): No debe haber puntos con intensidades mayores que tres veces la desviación estándar de ruido.

Criterio 1 (de mejora): En vista de que la sobredetección en la zona alta del espectro se mantenía aun aplicando el filtro de anchura, se decide tratar por separado la zona baja y la zona alta del espectro. La corrección de la línea de base en la región más elevada no logra la eficacia de la zona baja, por lo que una cierta curvatura residual se ve promocionada al multiplicar los espectros. Esta tendencia da lugar a una zona muy ruidosa que presenta dificultades en la identificación. Para mejorar este aspecto se trata por separado la zona inferior a 2300cm⁻¹ y la zona superior. Al tratar ambas zonas por separado, la eliminación de la línea de base puede llevarse a cabo de forma más específica en la zona alta, de manera que no perjudique a los picos inferiores, pero sí evite la promoción problemática.

Criterio 2 (de mejora): Para hacer el método más robusto, se decide imponer un criterio de anchura para los picos. Para ello se establece que han de contar con al menos tres puntos contiguos para considerarse acertados. Esta decisión se justifica con que la emisión Raman da lugar a picos anchos, cuya anchura disminuye, pero se mantiene, cuando se multiplican los espectros individuales. Al mismo tiempo, y al haber separado la zona baja de la alta del espectro, se establece un criterio de anchura más restrictivo (requiere picos más anchos) para la zona alta, ya que esta zona contiene las señales características de hidratación, o bandas del agua, y las vibraciones de los OH. Estas señales son, por lo general, más anchas que otras vibraciones presentes en la parte baja del espectro.

Criterio 3 (de mejora): También se restringe el número máximo de iteraciones a 16, pues en los espectros problema no se detectaban picos acertados adicionales a partir de esta iteración. Un estudio de este parámetro fue llevado a cabo estableciendo que entre 16 y 20 iteraciones los resultados que se obtienen son buenos. En caso de trabajar con nuevos espectros se sugiere que este parámetro sea susceptible de modificación dado que si las características espectrales son muy distintas podría ser un parámetro clave a la hora de evitar falsos positivos o ampliar la detección.

Criterio 4 (de parada): Como criterio último, se establece que cuando en una nueva iteración se detecten 9 picos o más, el método habrá entrado en la zona de ruido. Es extraño, y no se ha visto en ninguno de los espectros tratados, que para unas intensidades similares aparezcan tantos picos Raman. Es posible que en el caso de mezclas binarias, ternarias o más complejas esta situación sí que se dé y que en ese caso el parámetro resulte demasiado restrictivo, pero en todo caso lo que se busca es optimizar el método, pero minimizando el número de falsos positivos.

El método final, por lo tanto, tiene dos formas de finalizar la búsqueda: cuando detecta demasiados picos por iteración (criterio 4) o cuando haya descendido lo suficiente en intensidad como para considerar que la detección carece de interés (criterio 0). Este segundo criterio guarda una dependencia estrecha con las características del espectro bajo estudio.

Evaluación de las prestaciones del algoritmo de detección automática de picos Raman

A continuación se presentan resultados comparativos de los picos detectados automáticamente para el caso de la acumulación por multiplicación y para el promedio aritmético. Los datos de detección mediante promedio aritmético se obtienen usando el software SpectPro, que implementa una rutina de detección automática de picos utilizada en el desarrollo de RLS.

También se muestran las imágenes de los espectros de las distintas muestras con los picos detectados automáticamente por multiplicación. Habida cuenta de que los espectros multiplicados no permiten una visualización cómoda, se ha efectuado la raíz n-ésima para dicha representación.

Indeno En el caso del indeno (Figuras 46. y 47.) todos los picos detectados mediante la multiplicación de espectros coinciden con la bibliografía. No se han detectado falsos positivos. En este caso la detección basada en el promedio aritmético está sobredetectando los picos.

Indeno			
Multiplicación		Promedio aritmético	
Rshift (cm-1)	Intensidad (a.u.)	Rshift (cm-1)	Intensidad (a.u.)
739	233.9		
1023	367	1023	372.466
1210	401.1	1210	403.025
1555	536.9	1555	537.549
1612	455.7	1612	457.093
2889	431	2889	432.941
		2902	380.68
		3051	472.706
3065	566.5	3065	568.839
		3079	493.771
		3117	340.332

Figura 46: Comparación de picos detectados automáticamente mediante la multiplicación de espectros usando el algoritmo creado y mediante el promedio aritmético usando SpectPro. Ejemplo: indeno.

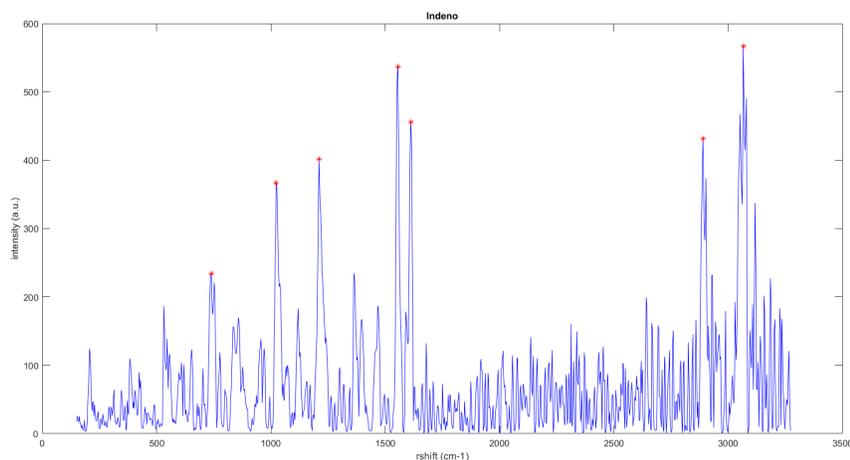


Figura 47: Detección automática de picos mediante multiplicación espectral para el caso del Indeno.

Calcita En este caso (Figuras 48. y 49.), el pequeño pico detectado para la calcita en 629cm^{-1} no se corresponde con un pico característico de este material. Sin embargo, sí que puede decirse que el espectro presenta una cierta señal en esa región. La detección de este pico se encuentra en el límite entre lo que consideramos acertado, aunque no sea lo esperado, o el punto en el que nos acercamos peligrosamente al ruido. Los dos pequeños picos situados a la derecha del pico principal presentan un comportamiento parecido y no son detectados. Esto demuestra que en este caso nos hallamos ante el límite de la detección.

Calcita			
Multiplicación		Promedio aritmético	
Rshift (cm-1)	Intensidad (a.u.)	Rshift (cm-1)	Intensidad (a.u.)
287	39699	286.652	39700
629	944		
721	4942		
1092	44029	1091.71	44029.3

Figura 48: Comparación de picos detectados automáticamente mediante la multiplicación de espectros usando el algoritmo creado y mediante el promedio aritmético usando SpectPro. Ejemplo: calcita.

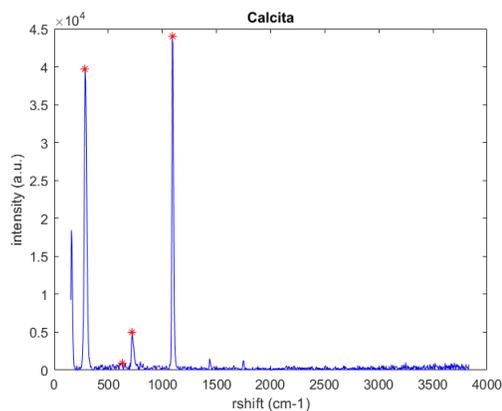


Figura 49: Detección automática de picos mediante multiplicación espectral para el caso de la calcita.

Ciclohexano De nuevo, los picos detectados en el caso del ciclohexano (Figuras 50. y 51.) se corresponden con aquellos presentes en la bibliografía. En este caso, el hecho de no ser capaces de detectar dobletes perjudica gravemente la detección. Sería idóneo ser capaces de descomponer la información espectral de la parte alta en los diversos picos presentes.

Ciclohexano			
Multiplicación		Promedio aritmético	
Rshift (cm-1)	Intensidad (a.u.)	Rshift (cm-1)	Intensidad (a.u.)
432	1756		
807	16877	806.974	16877.5
1037	7166		
1273	5994		
1450	5957		
2852	44931	2852	44931.3
		2937.97	39270.4

Figura 50: Comparación de picos detectados automáticamente mediante la multiplicación de espectros usando el algoritmo creado y mediante el promedio aritmético usando SpectPro. Ejemplo: ciclohexano.

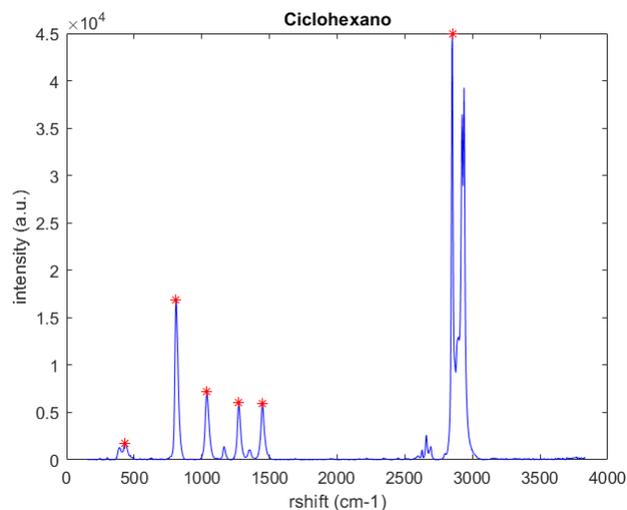


Figura 51: Detección automática de picos mediante multiplicación espectral para el caso del ciclohexano.

Hidróxido de Aluminio En el caso del hidróxido de aluminio (Figuras 52. y 53.) los picos de la zona alta sí que quedan separados en el espectro multiplicación, al menos hasta el límite de intensidades estudiado. Esto permite que se detecten por separado proporcionando gran riqueza en cuanto a la información espectral que se obtiene de la zona. En la parte baja del espectro aparece una zona abultada donde los picos de 741cm^{-1} y 901cm^{-1} no guardan una correspondencia adecuada con la bibliografía. Este tipo de señal es inusual y no ha sido considerado con detenimiento en la creación del algoritmo. Es muy importante notificar que en casos como este el método puede conducir a resultados erróneos. El balance global de la detección es positivo, pues se detectan numerosos picos más que en el caso de usar SpectPro.

AlOH3			
Multiplicación		Promedio aritmético	
Rshift (cm-1)	Intensidad (a.u.)	Rshift (cm-1)	Intensidad (a.u.)
328	8728		
551	17020	551	17019.8
741	839		
901	2660		
1017	1297		
3382	11132		
3446	26574	3446	26574.5
3578	19086	3578	19085.7
3699	11830		

Figura 52: Comparación de picos detectados automáticamente mediante la multiplicación de espectros usando el algoritmo creado y mediante el promedio aritmético usando SpectPro. Ejemplo: hidróxido de aluminio.

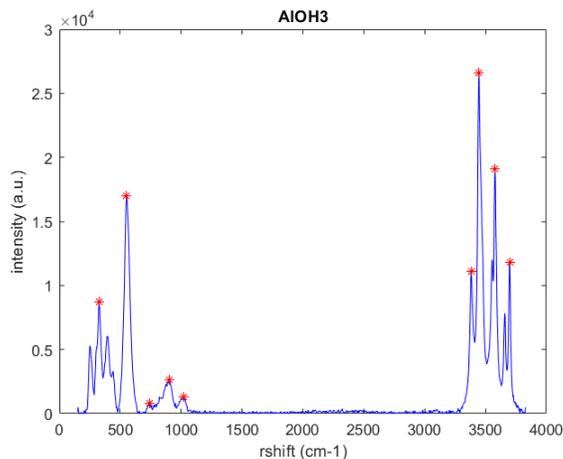


Figura 53: Detección automática de picos mediando multiplicación espectral para el caso del hidróxido de aluminio.

Talco En el caso del talco (Figuras 54. y 55.), se detectan dos picos: 471cm^{-1} y 1058cm^{-1} que no quedan justificados en la bibliografía. Sin embargo, al observar el espectro, se puede ver que existe señal en esa zona. De nuevo nos hallamos en el límite de detección al igual que sucedía en el caso de la calcita.

Talco			
Multiplicación		Promedio aritmético	
Rshift (cm-1)	Intensidad (a.u.)	Rshift (cm-1)	Intensidad (a.u.)
198	20596	198	20596.3
298	1159		
366	4098		
471	1295		
686	8684	686	8684.72
802	909		
1058	1134		
3724	7546	3724	7546.24

Figura 54: Comparación de picos detectados automáticamente mediante la multiplicación de espectros usando el algoritmo creado y mediante el promedio aritmético usando SpectPro. Ejemplo: talco.

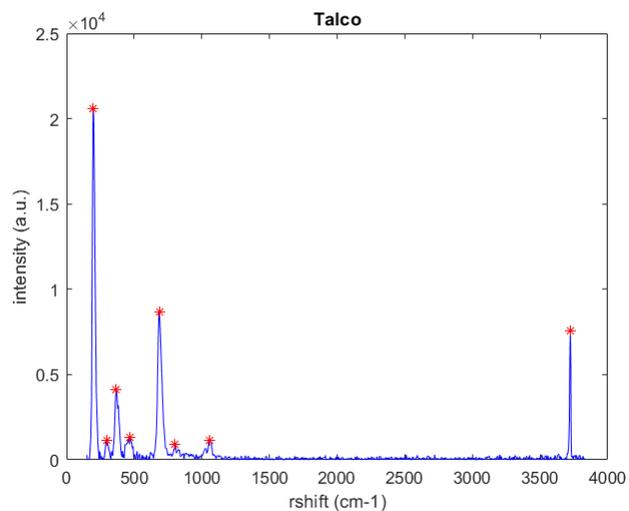


Figura 55: Detección automática de picos mediante multiplicación espectral para el caso del talco.

Vermiculita Es muy destacable el caso de la vermiculita (Figuras 56. y 57.), en el que aun tratándose de un espectro ruidoso logra mejorarse la detección. El pico secundario que identifica el método automático basado en la multiplicación de espectros proporciona una información complementaria al pico principal que será muy útil en el caso de identificar automáticamente muestras problema mediante comparación.

Vermiculita			
		Promedio aritmético	
Rshift (cm-1)	Intensidad (a.u.)	Rshift (cm-1)	Intensidad (a.u.)
196	14887	196	1497.85
683	3769		

Figura 56: Comparación de picos detectados automáticamente mediante la multiplicación de espectros usando el algoritmo creado y mediante el promedio aritmético usando SpectPro. Ejemplo: vermiculita.

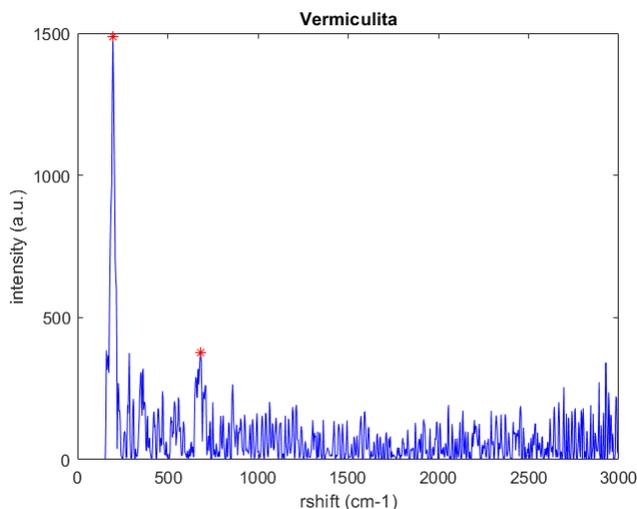


Figura 57: Detección automática de picos mediante multiplicación espectral para el caso del hidróxido de la vermiculita.

Puede observarse que gracias a las mejoras implementadas, se ha evitado la detección de falsos positivos y se ha logrado mejorar los resultados que se obtienen de forma automática mediante el promediado aritmético.

A la par, cabe destacar que frente a muestras con comportamientos diferentes en lo referente a su calidad espectral, el procedimiento ofrece buenos resultados indiscriminadamente. Por contar con unos resultados suficientemente poblados en cuanto variabilidad, no se han incluido los datos para la azurita, el hidróxido de Ca, el tetracloruro de Ca y el diamante. Los resultados presentados son suficientemente significativos y el comportamiento de las muestras restantes queda igualmente recogido en los ejemplos seleccionados.

En la búsqueda de perfeccionar el método, se llevó a cabo un estudio sobre como detectar dobletes. Ciertos picos que se unen en la base por lo que se detectan en conjunto y solo se identifica como pico

Raman el máximo de ambos. Esto supone un problema para ciertas muestras, como por ejemplo el caso del ciclohexano (véase Figura 51.), en el cual la no detección del doblete de la zona alta del espectro ($2800 - 3000\text{cm}^{-1}$). Se llevaron a cabo diversos intentos de mejorar este aspecto, pero ninguno ofreció resultados satisfactorios.

3.4. IDENTIFICACIÓN AUTOMÁTICA DE MINERALES

Llegados a este punto, se aborda el objetivo final del presente trabajo, que consiste en lograr la máxima automatización del proceso de identificación de materiales en función de su señal Raman. A tal efecto, se plantea concatenar la rutina de detección automática de picos explicada en las secciones anteriores con la algoritmia desarrollada por M. Khadmaoui (3.1.4), de forma que se obtenga un proceso totalmente automatizado para la identificación de materiales en base a su señal Raman, utilizando como entrada única el espectro en bruto sin ningún procesado previo.

El trabajo de integración de ambas rutinas ha requerido un esfuerzo de adaptación y optimización del algoritmo de correlación espectral de M. Khadmaoui para mejorar su comportamiento o resolver problemas detectados durante el análisis realizado en este trabajo.

Para la mejora de la detección se ha contribuido también a la ampliación de la base de datos mediante la adquisición de diversos espectros Raman, tal y como se recoge en el apartado 2.3.

3.4.1. Adaptación y optimización del algoritmo de clasificación de espectros

El algoritmo final sigue la misma estructura básica que el propuesto en trabajos anteriores (ver sección 3.1.4), con la excepción de que se introduce un paso adicional para la extracción de la información espectral del espectro problema en base a los algoritmos desarrollados previamente. De esta forma se obtiene la información de picos del espectro problema ($\langle bandPosition \rangle$ y $\langle bandIntensity \rangle$) de forma totalmente automática.

Adicionalmente, en este apartado se recogen las diversas modificaciones y mejoras realizadas sobre el algoritmo existente definido en la sección 3.1.4. Al haber realizado un trabajo preliminar de población de la base de datos, así como de utilización de muestras problema consideradas relevantes para el estudio de Marte, la casuística observada se ha incrementado muy apreciablemente con respecto a previos trabajos. Por tanto, se han detectado situaciones indeseables que se han corregido como se describe a continuación.

Rango espectral La primera modificación está relacionada con el rango espectral de los datos. Por un lado, se limita el ancho espectral al rango con el que trabajamos en los espectros ($150 - 3830\text{cm}^{-1}$). Este rango ha sido seleccionando atendiendo a la curvatura característica de los espectros, ya que es el que permite una correcta eliminación de la línea de base tras la división entre el borato, según se describe en la sección 3.2..

Por otro lado, se restringe la información de la base de datos al mismo rango que el espectro problema, debido a que la selección del material tiene en cuenta la cantidad de picos que coinciden en ambos. Esta coincidencia se vería negativamente afectada al comparar espectros de rangos distintos.

Actualización del cálculo de coincidencia y discriminación de espectros en base al estudio de las características Raman de diferentes fases y grupos minerales Se realizó un estudio comprensivo de las muestras y características espectrales de diferentes materiales y grupos de minerales. De esta forma, la obtención del subconjunto de espectros de la base de datos que pueden ser considerados coincidentes (paso 2 del algoritmo de clasificación) se pudo realizar teniendo en cuenta las características reales de los espectros, su variabilidad debida a diferentes composiciones y la información mineral que contienen. También se aplicó este estudio para actualizar la información espectral al final de cada iteración (paso 5 del algoritmo original en 3.1.4).

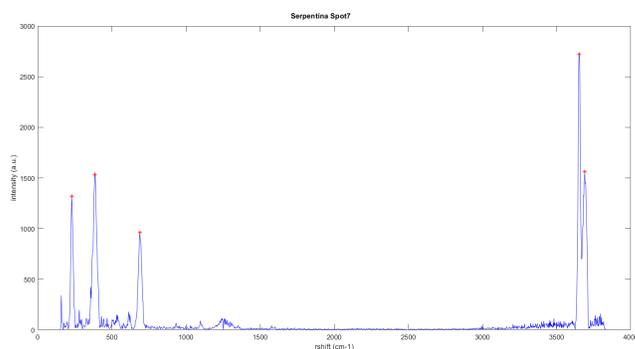
En concreto, en este estudio se realizan las modificaciones atendiendo a tres aspectos principales:

1- Identificación de las señales de hidratación en la muestra y procesamiento diferenciado según la región espectral. En primer lugar se estudia la presencia o no de hidratación en la muestra problema. Para ello se observa si el espectro tiene señales en la zona de los OHs. Esta condición tiene como objetivo incluir en la lista de posibles candidatos todas las muestras con hidrataciones semejantes presentes en la base de datos. Para ello, se redefine con un criterio más laxo, la tolerancia en posición (parámetro *tolerancia_posicion*) en el caso de muestras hidratadas para la búsqueda de candidatos. El valor estricto se recupera para el cálculo de coincidencias. Esta alteración del algoritmo original tiene dos ventajas: por un lado elimina la problemática asociada a una mayor presencia de ruido y falsos positivos en la zona alta del espectro (debida a la corrección por el borato de zinc así como de la línea de base, tal y como se discutió en secciones anteriores); y por otro promociona a los materiales hidratados de la base de datos para que puedan ser candidatos, habida cuenta de que la, en general, mayor anchura de los picos en la zona de los OHs da lugar a mayores inexactitudes en cuanto a la incertidumbre en posición de los picos de la zona alta.

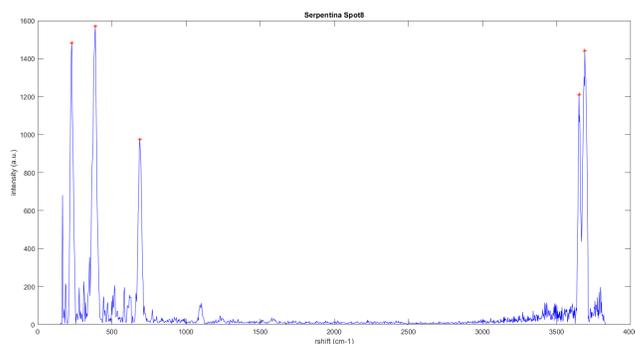
2- Caracterización de la variabilidad esperable en las posiciones/intensidades de los picos principales de algunas muestras Se ha observado en varias ocasiones que un mismo compuesto presenta variabilidad en su pico principal en función del punto de la muestra analizado. Esto es especialmente observable en aquellos materiales que presentan vibraciones con intensidades similares (por ejemplo, asociadas a la vibración bending de cationes diferentes en muestras con varios cationes en la estructura mineral). Esto se ha observado por ejemplo en espectros de serpentina, caolinita y vermiculita, entre otros. El ejemplo de la serpentina puede observarse en la Figura 58..

Es por ello, que el filtro de pico principal empleado en el algoritmo original de clasificación resulta excesivamente restrictivo. Lo que aquí se sugiere es comparar todos los picos de ambos espectros que posean una intensidad igual o superior a 0,6 (recordemos que los espectros están normalizados a 1) y observar si alguno de ellos coincide permitiendo 9cm⁻¹ de variabilidad entre la posición de ambos. Si se produce la coincidencia de alguno de los picos, la muestra de la base de datos pasa a considerarse un posible candidato como mineral de la muestra problema.

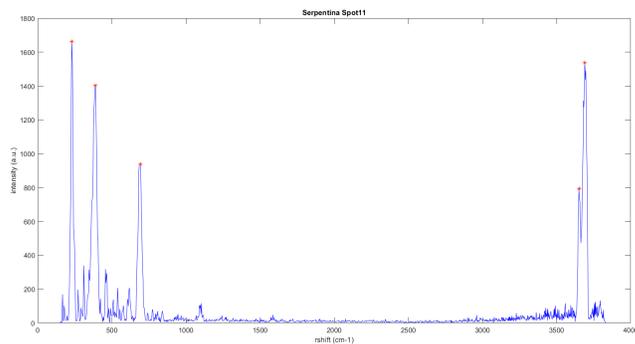
Variabilidad de picos principales de la serpentina



(a) Punto número 7 de la muestra de serpentina con pico principal en $3650,82\text{cm}^{-1}$.



(b) Punto número 8 de la muestra de serpentina con pico principal en $388,31\text{cm}^{-1}$.



(c) Punto número 11 de la muestra de serpentina con pico principal en $229,866\text{cm}^{-1}$.

Figura 58: Variabilidad en el pico principal de la serpentina.

Los espectros mostrados en la Figura 58. pertenecen a distintos puntos de la misma muestra. Como puede observarse, el pico principal es distinto en cada caso: en Spot7 se sitúa en $3650,82\text{cm}^{-1}$ (Figura 58a.), en Spot8, en $388,31\text{cm}^{-1}$ (Figura 58b.) y en Spot11 en $229,866\text{cm}^{-1}$ (Figura 58c.). Esta variabilidad puede deberse a que la orientación de los granos minerales puede variar, lo cual implica una variación en intensidad tal y como se recoge en la sección 1.3.1.

3- Discriminación y descarte en base a la pertenencia a un mismo grupo mineral. Una vez encontrado el mejor candidato (paso 4 del algoritmo), se verifica si éste forma parte de un grupo mineral, tal y como ocurre con la serpentina o la vermiculita. De ser así, el proceso de eliminación de las bandas identificadas no elimina únicamente dichas bandas, sino que además se eliminan todos los picos coincidentes de los diversos espectros de dicho grupo presentes en la base de datos. Esta modificación sirve para restar todos los picos de dicho grupo mineral y no dejar alguno susceptible de ser identificado como otro material. El rationale detrás de esta mejora es que la variabilidad en intensidad de diversos picos de la serpentina puso de manifiesto que no eran eliminados en la primera iteración de la identificación y que por ello quedaban formando parte del espectro problema para ser nuevamente identificados. Dado que el espectro había perdido parte de los picos de dicho material la coincidencia no era buena y los candidatos propuestos para estos picos olvidados solían ser erróneos. Es decir, las posibles variabilidades en determinadas vibraciones de materiales con mineralogía compleja pueden hacer que los 9cm-1 de incertidumbre aceptada para considerar una coincidencia no sean suficientes como para dar respuesta a todas las posibles variaciones y desplazamientos que pueden producirse de forma natural en un espectro.

3.4.2. Definición del algoritmo mejorado para la detección y la clasificación automática

De este modo se define el algoritmo final capaz de realizar la clasificación de las muestras en base a su información espectral.

El algoritmo final comienza abriendo los espectros y cortándolos en un rango inicial entre 121cm^{-1} y 4000cm^{-1} , para después dividirlos entre el espectro de borato de zinc amorfo y cortarlos nuevamente entre 150cm^{-1} y 3830cm^{-1} . La división entre el borato, que toma valores más bajos de intensidad en sus extremos, deforma el espectro en estas regiones, por ello se efectúa un nuevo corte tras la división. En este punto, se separan el comportamiento de la parte baja del espectro (desplazamientos Raman menores a 3000cm^{-1}) y la parte alta (desplazamientos Raman mayores de 3000cm^{-1}). La eliminación de la línea de base se lleva a cabo por separado en ambas regiones, pues así se corrige mejor la curvatura de la zona alta del espectro, tal y como se explicó en la sección 3.3.4. Los espectros de cada región son entonces multiplicados y se someten por separado al algoritmo de detección automática de picos. El número de iteraciones permitidas y el criterio para el filtro de anchura son diferentes en cada zona. Sus valores se muestran en la Tabla 10.

Parte del código empleado en este proceso puede consultarse en el ANEXO (sección 6).

	Zonal baja	Zonal alta
Número de iteraciones	20	15
Puntos requeridos por el filtro de anchura	3	6

Tabla 10: Parámetros empleados en la detección automática para las diferentes regiones del espectro.

Una vez que se ha obtenido la información de picos de las dos regiones se fusionan los datos obtenidos y se procede a la identificación de materiales.

En este punto, se carga la base de datos disponible en MATLAB y se acota la información al rango espectral bajo estudio. Los valores de intensidad y desplazamiento Raman de las diversas muestras de la base de datos quedan disponibles en vectores para su cómoda comparación.

A continuación, mediante un condicional se establece si la muestra está o no hidratada. Si lo está, lo deseable es incluir como candidatos todos los espectros que cuenten con señales Raman de hidratación semejantes a la muestra problema. Para incluir la variabilidad de posición en esta región se amplía el parámetro *tolerancia_posición* a 100cm^{-1} , para el caso de espectros hidratados, así la propuesta de candidatos es mucho más permisiva.

Con el fin de considerar el comportamiento real de los espectros, se comparan los picos más intensos (intensidad igual o superior a 0.6) tanto de la muestra problema, como de las muestras de la base de datos. Si se encuentra algún pico que coincida en posición dentro de este margen de intensidad, se selecciona la muestra de la base de datos como posible candidata.

Ser más tolerantes en un comienzo no supone un problema, ya que para el cálculo de la coincidencia (ecuación 22) retomamos el valor de *tolerancia_posición* = 9cm^{-1} . De este modo, el candidato final propuesto como material presente en la muestra ha debido cumplir criterios de coincidencia estrictos.

En el momento en el que se selecciona un candidato final, se comprueba si pertenece o no a un grupo mineral. Si la respuesta es afirmativa, como sucede en el caso de las serpentinas y las vermiculitas, se eliminan del espectro problema todas las señales que coincidan con picos de alguna de las muestras pertenecientes al grupo mineral. De ser negativa, se resta exclusivamente la contribución del candidato propuesto.

Una vez que se han eliminado los picos característicos del primer material, se usa la información de picos restante para la búsqueda un posible material secundario. Este comportamiento iterativo se detiene cuando todos los picos han sido asignados a señales de la base de datos, o cuando en dos iteraciones consecutivas no se resta ningún pico.

Este método en su conjunto, pretende restar impacto a los desplazamientos que sufran los picos de los espectros si la muestra está hidratada, así como abarcar las variabilidades propias de los grupos minerales. Se probaron metodologías mucho más tolerantes y también más estrictas que introdujeron errores en la clasificación. Con ello, se ha llegado al compromiso expuesto, con el que se obtiene un porcentaje de aciertos significativo, tal y como se presenta a continuación.

Para visualizar de forma esquemática lo expuesto en el presente apartado consúltese la Figura 59.

Detección automática de picos e identificación de materiales

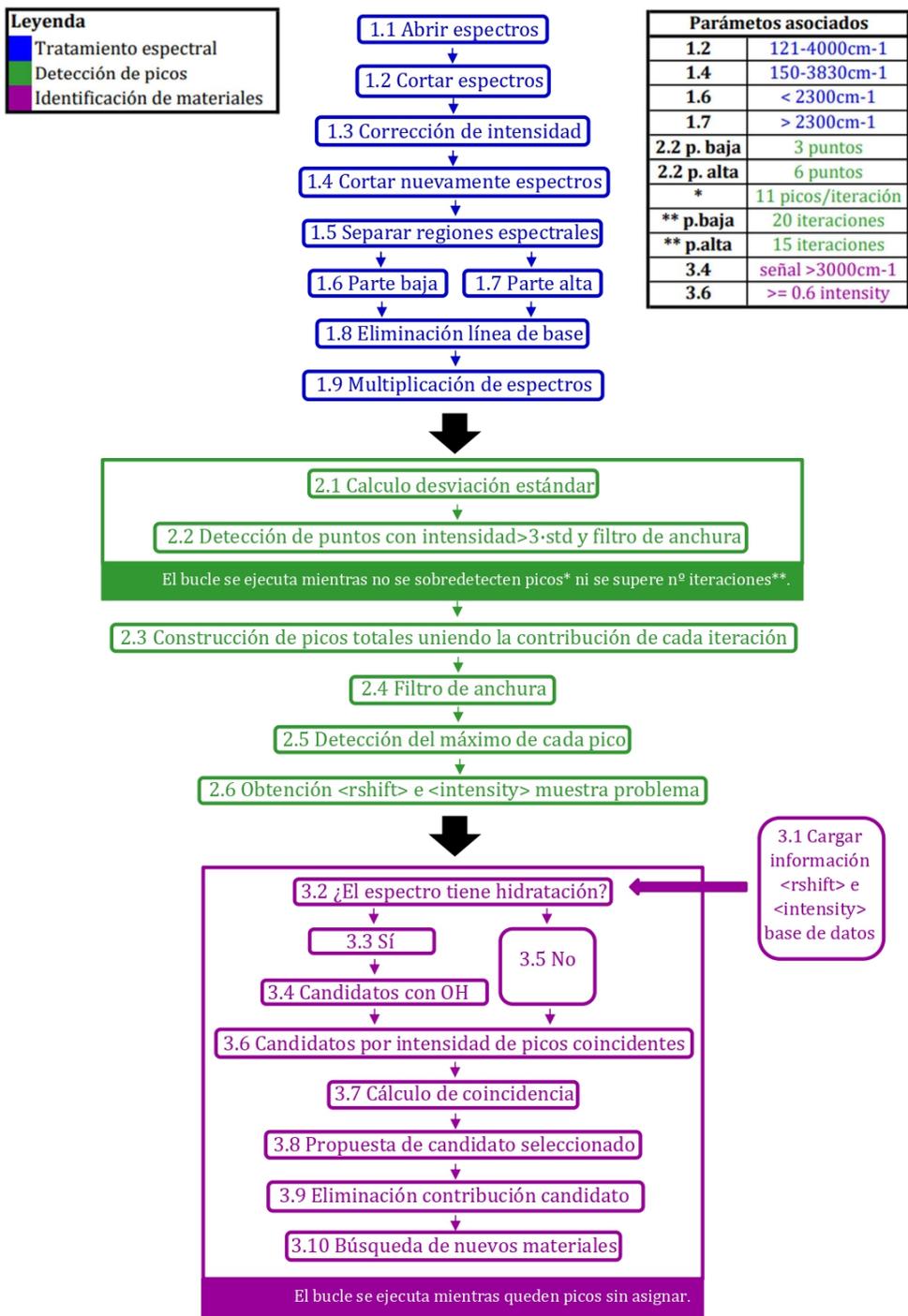


Figura 59: Diagrama de los pasos seguidos por el código completo de detección de picos e identificación de materiales.

Evaluación de prestaciones del algoritmo de clasificación totalmente automático A continuación se recogen los resultados obtenidos mediante el algoritmo explicado en la sección 3.4.2.

Empleando el instrumento RLS con el cabezal BWTEK-2008 (“i-Raman”) se han obtenido los resultados que se muestran en la Figura 60., donde el color verde significa una detección acertada y el rojo una detección errónea.

	Spot	Primer candidato	Segundo candidato	Tercer candidato
Olivino	0	fosterita	caolinita	
	1	fosterita		
	2	fosterita	serpentina (antigorita)	
	3	fosterita		
	4	fosterita	grafito	
Serpentina	5	fosterita		
	6	serpentina (INDURAM)		
	7	serpentina (antigorita)		
	8	serpentina (INDURAM)		
	9	serpentina (INDURAM)		
	10	serpentina (lizardita)		
Vermiculita	11	serpentina (INDURAM)		
	0	vermiculita	pirolusita	
	1	vermiculita	piroxeno	
	2	vermiculita	piroxeno	
	3	vermiculita	pirolusita	goetita
	4	vermiculita		
Caolinita	5	vermiculita	pirolusita	
	12	caolinita	anatasa	
	13	caolinita		
	14	caolinita	goetita	
	15	caolinita		
	16	caolinita	anatasa	
Serpentina-Olivino 47,5	17	caolinita	goetita	
	36	serpentina (lizardita)	fosterita	
	37	serpentina (INDURAM)	fosterita	
	38	serpentina (lizardita)	fosterita	
	39	serpentina (lizardita)	fosterita	
Vermiculita-Kaolinita 50	40	serpentina	fosterita	
	41	serpentina (lizardita)	fosterita	
	6	caolinita		
	7	caolinita		
	8	caolinita	esfalerita	
	9	caolinita	serpentina (lizardita)	
10	caolinita	esfalerita		
11	caolinita	esfalerita		

Figura 60: Resultados de la identificación automática para el instrumento RLS con el cabezal BWTEK-2008 (“i-Raman”)

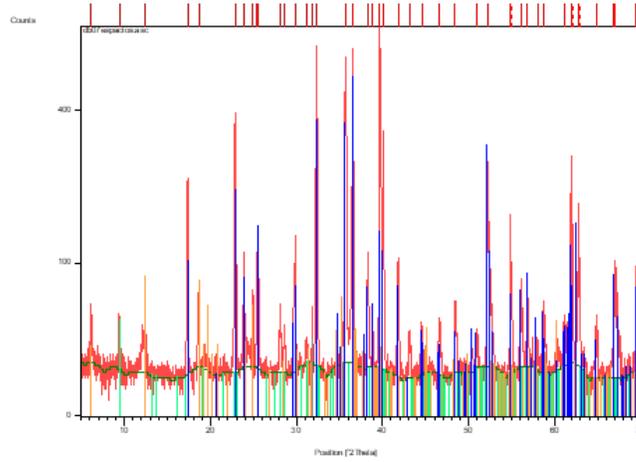
El primer candidato ofrece un resultado correcto en el 100 % de los casos. Mientras que para el segundo candidato solo ofrece aciertos en el 50 % de los casos. El tercer material propuesto es un error. Se debe especificar que la no propuesta de un material erróneo es considerada como acierto.

Con este estudio, puede verse que se han igualado los resultados obtenidos por M. Khadmaoui presentados en el apartado 3.1.4 en el caso de la primera iteración. En el caso de la segunda iteración el porcentaje de falsos positivos es ligeramente mayor, un 44 %. A la par, por no detectar un segundo material en dos de los puntos de la muestra de vermiculita y caolinita, se considera que el método infravalora los materiales en un 6 % de los casos para la segunda iteración.

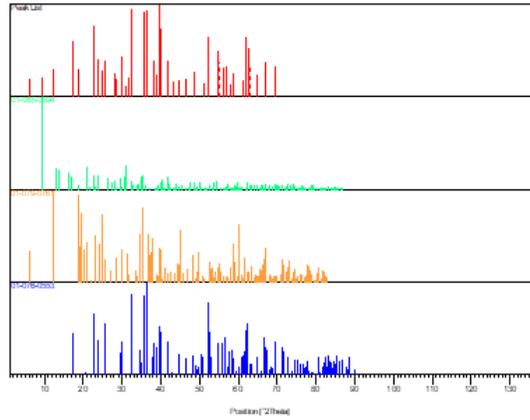
En el caso de los falsos positivos propuestos en la muestra de olivino, no pueden considerarse completamente erróneos, puesto que un estudio más detallado de esta muestra mediante difracción de rayos-X, puso de manifiesto que, a pesar de haber utilizado un estándar geológico certificado, en su composición cuenta con un 20 % de clinocloro. Los resultados de este estudio pueden encontrarse en la Figura 61. En este caso, el algoritmo fue capaz de detectar la presencia de un segundo material en la muestra, pero no pudo acertar en su naturaleza, pues no se cuenta con una referencia de clinocloro en la base de datos. Si se considera que el segundo material debía ser detectado, el porcentaje de falsos positivos se reduce al 36 %.

Se ha demostrado que la agrupación de diferentes variedades de un mismo material, para realizar de forma conjunta la eliminación de picos en el proceso iterativo, es un acierto a la hora de mejorar el comportamiento del algoritmo. Por ejemplo, en el caso de la serpentina, al agrupar sus diferentes variedades (crisotilo, lizardita y antigorita), se eliminaron los falsos positivos previamente propuestos.

A su vez, los resultados erróneos para el segundo y tercer candidato de la vermiculita se justifican con que la fórmula química de este mineral es altamente variable. Bajo la denominación de vermiculita se recogen diversos minerales cuya variabilidad dificulta la detección. La tolerancia en posición establecida (9cm^{-1}) no es suficiente como para acoger las variaciones descritas y no todos los picos son identificados en la primera iteración, dando lugar a propuestas erróneas en las siguientes. Este problema podría solucionarse poblando más densamente la base de datos (véase FUTURAS LÍNEAS DE MEJORA, apartado 5.).



(a) Difractograma de la muestra problema.



(b) Difractograma separado en función de sus componentes.

Ref. Code	Chemical Formula	Score	Compound Name	RIR	SemiQuant [%]
01-085-0594	Ca _{1.95} Al _{3.9} Si _{8.1} O ₂₄	50	Chabazite	8,980	<1
01-079-0761	(Mg _{4.715} Al _{6.94} Fe _{2.69} Fe _{1.09} Cr _{1.28} Ni _{0.11}) (Si _{3.056} Al _{1.944})O ₁₀ (OH) ₈	26	Clinocllore	0,700	~20
01-076-0553	Mg _{1.460} Fe _{0.540} Si O ₄	67	Forsterite	0,980	~80

(c) Composición de la muestra problema.

Figura 61: Estudio de difracción de Rayos-X para la muestra de olivino con clinocllore.

3.5. CONCLUSIONES DEL TRABAJO COMPUTACIONAL

Se puede concluir, en base a los resultados obtenidos, que se ha logrado mejorar la detección automática implementada en la herramienta SpectPro, basada en espectros promediados aritméticamente. La ventaja que ofrece la multiplicación de espectros ha quedado demostrada a lo largo del trabajo, sin embargo, no sería justo atribuirle todo al método de acumulación por multiplicación.

Tras un exhaustivo trabajo de entendimiento del comportamiento de los espectros, de su línea de base, de la promoción del ruido en las diversas zonas, de la anchura de la emisión Raman, de las intensidades relativas entre picos, etc., se ha implementado este método que recoge todas estas características. La contribución de todas ellas es lo que hace que el método funcione con la calidad demostrada. Dejar de considerar alguna característica o intentar que sean eficaces por separado dará lugar a resultados mucho más pobres. En definitiva, se puede concluir que se ha logrado entender cómo la información Raman se pone de manifiesto en un espectro real, tomado por un equipo de verdad, como los que cualquier experimentador manejará en su vida científica. Este entendimiento, ha permitido crear un algoritmo que considere todos estos aspectos y sugiera, con un acierto notable, donde se encuentra la información Raman.

Cabe mencionar, y así el lector puede haberlo notificado, que los picos Raman que han sido detectados a lo largo de esta parte del trabajo están ligeramente desplazados respecto a las posiciones propuestas en la bibliografía. Este desplazamiento se debe al comportamiento del instrumento FS, que durante la campaña de adquisición de las muestras aquí utilizadas estaba ligeramente descalibrado.

En lo referente a la identificación automática de materiales, las mejoras llevadas a cabo han logrado igualar el porcentaje de aciertos que el procedimiento previamente existente, con la diferencia de que en este caso se concatenan algoritmos totalmente automáticos. El algoritmo primitivo, aunque muy robusto computacionalmente, no consideraba las características propias que tienen los minerales debido a su naturaleza. La complejidad de estos minerales y sus variaciones, así como las consecuencias que estas tienen en espectroscopia Raman, han hecho necesario un estudio detallado acerca de los nuevos requerimientos para el algoritmo. Así surgió la idea de sustituir el filtro de pico principal por un filtro mucho más tolerante que considerara un margen amplio de intensidades. De igual modo, se quiso destacar la importancia de la hidratación en la selección de candidatos. También, debido a las variaciones compositivas que dan lugar a desplazamientos Raman, se implementó la agrupación de minerales en grupos para la resta de materiales.

4. CONCLUSIONES FINALES

La detección automática de minerales se presenta como una herramienta de gran utilidad para dar soporte a la misión ExoMars. Cuando lleguen a la Tierra espectros Raman procedentes de Marte, podrán ser sometidos al algoritmo de detección automática de picos e identificación de materiales. Con ello, el experimentador obtendrá una primera propuesta acerca de que material ha sido analizado.

La idea que motivó el desarrollo del presente trabajo fue explorar la multiplicación de espectros, que es la base del algoritmo de detección. Pese a que en primera instancia la visualización del espectro multiplicación no ofrece una ventaja aparente, el incremento espectacular de la SNR sí que lo hace, facilitando la detección automática de picos.

Para lograr aprovechar los beneficios de la multiplicación se exploraron numerosas vías, de las cuales solo se ha dado cuenta parcialmente. Estudios adicionales sobre representaciones logarítmicas y con raíces n -ésimas también fueron llevados a cabo, así como de dependencia de la relación señal-ruido con el número de acumulaciones. Tras todos los intentos puestos en práctica, se puede concluir que se ha llegado a un algoritmo que detecta con acierto los picos Raman de espectros creados mediante acumulación por multiplicación. El éxito demostrado motiva la utilización del algoritmo en casos en los que los espectros sean difíciles de interpretar. Se ha visto que en espectros ruidosos, por ejemplo en el caso del indeno (véase Figura 47.), la detección de picos es bastante selectiva y suele evitar los falsos positivos.

Al unir la detección de picos a la algoritmia de correlación espectral, la identificación de un material en base a su señal Raman pasa a ser un método completamente automatizado, lo cual supone una ventaja significativa, que además consigue resultados similares a los algoritmos propuestos en trabajos anteriores, donde el procesamiento de los espectros era realizado de forma manual por un operador.

Bien es cierto, que el método aún ofrece un amplio margen de mejora, especialmente en lo que concierne a la clasificación de minerales. La complejidad de las soluciones sólidas introduce la dificultad de tener que considerar diversas fórmulas minerales bajo la denominación de un mismo compuesto. Las variaciones compositivas se traducen en variabilidad espectral. El *modus operandi* actual aún no está preparado para considerar al completo estas diferencias.

Sin embargo, las modificaciones introducidas en el algoritmo de identificación han logrado resolver casos problemáticos como el de la antigorita (sección 3.4.2) y la ampliación de la base de datos ADAMM ha proporcionado soluciones para casos como el de la caolinita, ya que al contar con una base de datos más densamente poblada, la comparación de compuestos es más exhaustiva.

Cabe mencionar de forma especial, lo mucho que se ha aprendido acerca de la calibración de instrumentos. Al introducir diversos equipos en la realización del trabajo, ha sido necesario familiarizarse con todos ellos: con un espectrómetro estándar, como es el espectrómetro de doble pista y con un espectrómetro de nueva generación, como el simulador del instrumento RLS. A la par, la creación de SpectrumCalibrator supuso un reto computacional no enfrentado previamente.

En resumen, en este Trabajo de Fin de Grado se ha realizado un estudio que cubre aspectos teóricos, experimentales, de calibración y procesamiento de datos para lograr la detección de picos e identificación de minerales a partir de espectros. El conjunto del trabajo se enmarca en el contexto de la misión ExoMars, sin embargo, su proyección va más allá pues ha permitido obtener una visión global acerca de la espectroscopia Raman, la cual ha sido empleada para un caso real concreto, el instrumento RLS. Aún quedan por explorar muchas de las aplicaciones que puede tener este algoritmo y cuáles son

sus limitaciones. Bajo este fin, se han identificado a lo largo del estudio, una serie de líneas futuras que permitirían completar y mejorar los métodos expuestos si fuesen puestas en práctica.

5. FUTURAS LÍNEAS DE MEJORA

A lo largo del desarrollo de este trabajo, han surgido ideas complementarias a las explicadas previamente. Explorar estas nuevas vías sería muy interesante para el perfeccionamiento del método.

- Una de las propuestas relacionadas con la detección automática pasa por estudiar la multiplicación de espectros por grupos implementando el parámetro N_{av} (véase la sección 2.1.2.) en el código creado. De este modo, el procedimiento podría simular mejor el comportamiento del instrumento de vuelo y a la par se podría estudiar cual es la combinación que proporciona mejores resultados.
- La detección automática de picos se encuentra con serios problemas cuando el espectro cuenta con dobletes que se detectan como un mismo pico. Se trató de solucionar este conflicto, pero no fue posible. La detección de dobletes aportaría mucha información en el caso de espectros cuya característica principal sea la presencia de estos.
- Se ha visto que, para solventar la variabilidad del pico principal en la detección automática de minerales, muchas veces no es suficiente con aumentar el rango de intensidades considerado en la comparación. Este problema podría solucionarse ampliando la base de datos con varios espectros de distintos puntos de un mismo mineral.
- La promoción que sufren los picos al multiplicar varios espectros es significativa. Se ha visto también que al multiplicar un espectro por si mismo sus picos se promocionan (26). Tal vez esta vía facilitaría la detección automática en el caso de contar con un solo espectro.
- Se cuenta con muestras adicionales que incluyen piroxenos y feldepatos, que han sido adquiridas con el RLS Sim y un nuevo cabezal integrado (Spectra Solutions). Este cabezal cuenta con una función aparato difícil de corregir, ya que introduce un ruido muy elevado en el límite inferior del espectro. Poder analizar estas muestras, cuyo comportamiento es muy dispar al de aquellas estudiadas en este trabajo, permitiría reforzar el método. Se estima que este nuevo estudio debe ser una propuesta doble: por un lado, deben investigarse las posibles mejoras de hardware destinadas a reducir la contribución de la función aparato y por otro lado, debe llevarse a cabo una investigación exhaustiva de la corrección en intensidad y la eliminación de la línea de base para las nuevas muestras.
- Existen estudios (25) donde la clasificación dentro de un grupo mineral atiende a los picos Raman presentes en la muestra. Para ello se establecen condicionales que clasifican la muestra en subgrupos, hasta dar con el mineral que ofrece una mejor coincidencia. Este estudio sería especialmente interesante en el caso de piroxenos.

Referencias

- [1] A. G. Moral, 2019, *Desarrollo del instrumento RLS (Raman Laser Spectrometer) para la misión ExoMars: nuevos retos de la espectroscopia Raman en la exploración planetaria*, Universidad de Valladolid, Valladolid.
- [2] C. Quantin-Nataf, J. Carter, L. Mandon et al., 2021, *Oxia Planum: The Landing Site for the ExoMars “Rosalind Frankli” Rover Mission: Geological Context and Prelanding Interpretation*, Astrobiology 10.1089/ast.2019.2191.
- [3] M. H. Carr, J. W. Head III, 2010, *Geologic history of Mars*, Earth and Planetary Science, Volume 294, Issues 3–4, 1 June 2010, Pages 185-203.
- [4] F. Rull, S. Maurice, I. Hutcheon et al, 2017, *The Raman Laser Spectrometer for the ExoMars Rover Mission to Mars*, Astrobiology 10.1089/ast.2016.1567.
- [5] P. Vandenabeele, 2013, *Practical Raman Spectroscopy - An Introduction*, Universidad de Ghent, Bélgica, John Wiley & Sons Ltd. Print ISBN:9780470683194, Online ISBN:9781119961284.
- [6] F. Rull, comp., 1993, *Espectroscopia Infrarroja y Raman: Fundamentos y Aplicaciones en Mineralogía*, Valladolid, España, Universidad de Valladolid. ISBN:847762349X.
- [7] G. López-Reyes, 2015, *Development of algorithms and methodological analyses for the definition of the operation mode of the Raman Laser Spectrometer instrument*, Universidad de Valladolid, Valladolid.
- [8] A. Sanz-Arranz, J.A. Manrique-MARTINEZ, J. Medina-Garcia, et al., 2006, *Amorphous zinc borate as a simple standard for baseline correction in Raman spectra*, Journal of Raman Spectroscopy, DOI 10.1002/jrs.5144.
- [9] A. Arbor, 2002, HoloSpec™ Imaging Spectrograph. OPERATIONS MANUAL, Kaiser Optical Systems, Inc. 371 Parkland Plaza.
- [10] A. Arbor, 2002, Raman Calibration Accessory. OPERATIONS MANUAL 2004948 V3.0, Kaiser Optical Systems, Inc. 371 Parkland Plaza.
- [11] I-M. Chou, 2014, *Calibration of Raman Shifts of Cyclohexane for Quantitative Analyses of Methane in Natural and Synthetic Fluid Inclusions*, International GeoRaman Conference Sanya Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan 572000, China.
- [12] S.K. Sharma, P.G. Lucey, M. Ghosh, et al., 2003, *Stand-off Raman spectroscopic detection of minerals on planetary surfaces*. Spectrochimica Acta.
- [13] N.N. Greenwood y A. Earnshaw, 1997 *Chemistry of the Elements*, Capítulo 9: Silicon, página 357.
- [14] M. Veneranda, A. Sanz-Arranz, J. A. Manrique, et al, 2021, *Analytical database of Martian minerals (ADaMM): Project synopsis and Raman data overview*, Journal of Raman Spectroscopy.
- [15] G. López-Reyes, F. Rull Pérez, 2017, *A method for the automated Raman spectra acquisition*, Journal of Raman Spectroscopy, DOI 10.1002/jrs.5185.
- [16] M. Khadmaoui Bichouna, 2020, *Desarrollo de algoritmos para la identificación automática de espectros Raman de la misión ExoMars*, Universidad de Valladolid, Valladolid.

- [17] H. D. Ruan, R. L. Frost y J. T. Kloprogge, 2001, *Comparison of the Raman spectra of Bayerite, Boehmite, Diaspore and Gibbsite*, Journal of Raman Spectroscopy.
- [18] R. L. Frost, W. N. Martens, L. Rintoul, et al., 2002, *Raman spectroscopic study of azurite and malachite at 298 and 77 K*, Journal of Raman Spectroscopy.
- [19] D. Krishnamurti, 1957, *The Raman Spectrum of Calcite and its interpretation*, Indian Academy of Sciences.
- [20] Z. V. Padanyi, 1970, *The Raman Spectrum of Ca(OH)₂*, Solid State Communications.
- [21] R.L. McCreery, 2000, *Raman Spectroscopy for Chemical Analysis*.
- [22] M. Arab, D. Bougeard y K. S. Smirnov, 2002, *Experimental and computer simulation study of the vibrational spectra of vermiculite*, Physical Chemistry Chemical Physics.
- [23] B. Reynard, L. Bezacier y R. Caracas, 2015, *Serpentines, talc, chlorites, and their high-pressure phase transitions: a Raman spectroscopic study*, Physics and Chemistry of Minerals.
- [24] R. S. Krishnan, 1945, *Raman Spectrum of Diamond*, Nature.
- [25] A. Wang, B. L. Jolliff, L. A. Haskin et al, 2001, *Characterization and comparison of structural and compositional features of planetary quadrilateral pyroxenes by Raman Spectroscopy*, American Mineralogist
- [26] R. Pintelon, J. Schoukens y J. Renneboog, 1988, *The Geometric Mean Power (Amplitude) Spectra has a Much Smaller Bias Than the Classical Arithmetic (RMS) Averagin*, IEEE Transactions on instrumentation and measurement.

6. ANEXO

En este anexo se presenta el esquema del algoritmo creado para la identificación automática materiales en base a su señal Raman, así como el algoritmo modificado para la identificación de compuestos.

Desde *Automatic_peak_identification.m* se seleccionan las muestras a analizar y se llama al resto de scripts en el orden que sigue:

- 1º *A_open_and_cut_spectrum.m*
- 2º *B_BZn_divided_baseline_removed.m*
- 3º *c_multiplicacion_espectros.m*
- 4º *OH_c_multiplicacion_espectros.m*
- 5º *d_deteccion_automatizada_de_picos.m*
- 6º *OH_d_deteccion_automatizada_de_picos.m*
- 7º *B_normalizacion_tolerant_group_clasificacion.m*

Donde *B_normalizacion_tolerant_group_clasificacion.m* llama a *a_extract_bds.m* y a las funciones *tolerantPeak_coincidence_fn.m* y *coincidence_calculator_fn.m*

A su vez, los script *textitd_deteccion_automatizada_de_picos.m* y *OH_d_deteccion_automatizada_de_picos.m* llaman a las funciones *absoluteMaxima_detection_fn.m* y *maximaPeakDetection_fn.m* y *absoluteMaxima_detection_fn.m* y *maximaPeakDetection_fn.m*, respectivamente. Cada función incluye los parámetros característicos de la región del espectro que estudia, tal y como se recoge en la Tabla 10.

Como ejemplos significativos, se muestra el código del script *d_deteccion_automatizada_de_picos.m* y de la función *absoluteMaxima_detection_fn.m*.

6.1. d_deteccion_automatica_de_picos.m

```
nbOfIterations = 20;
%We have seen that this number of iterations is a good amount for the save detection of
peaks.
std_param = 3;
amount_of_peaks_thereshold = 11;
M = zeros(nbOfRows_L,nbOfIterations);
intensity_bzn_divided_baseline_removed_geom_avg_L =
abs(intensity_bzn_divided_baseline_removed_geom_avg_L);
standard_deviation = std(intensity_bzn_divided_baseline_removed_geom_avg_L);
%first value of std.
intensity_bzn_divided_baseline_removed_geom_avg_std_modified =
intensity_bzn_divided_baseline_removed_geom_avg_L;
%intensity to change the values of the detected peaks.

% Vectors to keep the values of the peaks and the iteration in which they
% have been detected.
rshift_gscatter = zeros(numel(intensity_bzn_divided_baseline_removed_geom_avg_L),1);
intensity_gscatter = zeros(numel(intensity_bzn_divided_baseline_removed_geom_avg_L),1);
intensity_peak_detection =
zeros(numel(intensity_bzn_divided_baseline_removed_geom_avg_L),nbOfIterations);
group_gscatter = zeros(numel(intensity_bzn_divided_baseline_removed_geom_avg_L),1);
indexes_vector = 0;
indexes_vector_peak_detection = 0;

g = 1;%iteration counter

for a=nbOfIterations:-1:1%this loop is useful when we change the value of standard_deviation
for j=numel(intensity_bzn_divided_baseline_removed_geom_avg_std_modified):-1:1
if intensity_bzn_divided_baseline_removed_geom_avg_std_modified(j)>
=(std_param*standard_deviation)
intensity_gscatter(j)=intensity_bzn_divided_baseline_removed_geom_avg_L(j);
intensity_peak_detection(j,a)=
intensity_bzn_divided_baseline_removed_geom_avg_L(j);
rshift_gscatter(j)=rshift_c_L(j);
group_gscatter(j) = g;
indexes_vector(j) = j;
indexes_vector_peak_detection(j,a) = j;
intensity_bzn_divided_baseline_removed_geom_avg_std_modified(j) = 0;
% We replace the intensity values which have been detected by zero.
end
end

% Maxima detection for representation
% We are detecting the maximum of every region in every iterarion.
M, true_M
= maximaPeakDetection_fn(intensity_bzn_divided_baseline_removed_L,indexes_vector,indexes_vector_peak_detec

non_zero_elements = find(true_M(:,a));
```

```

%If we detect more than 11 peaks during an iteration, the loop ends%because the detection
can be close to the noise signal. amount_of_peaks_of_every_iteration(:,g) = numel(non_zero_elements)-1;
if amount_of_peaks_of_every_iteration(:,g)>=amount_of_peaks_thereshold break end
g = g+1;
standard_deviation = std(intensity_bzn_divided_baseline_removed_geom_avg_std_modified);

end
% Absolute maxima detection for representation% This part of the script detects the
absolute maximum of every peak% region, it does not matter in what interaction this peak
has been% detected.
[M_absolute_maximum] =
absoluteMaxima_detection_fn(intensity_bzn_divided_baseline_removed_L,indexes_vector,intensity_gscatter);

sample_peaks = M_absolute_maximum;

non_zero_elements_sample_peaks = find(sample_peaks);
nbOfElements = numel(non_zero_elements_sample_peaks);

figure1 = figure;
plot(rshift_c_L,intensity_bzn_divided_baseline_removed_geom_avg_L.^(1\nbOfSpectra),'blue');
hold on
plot(rshift_c_L(non_zero_elements_sample_peaks),
sample_peaks(non_zero_elements_sample_peaks).^(1\nbOfSpectra),'r*'); hold on
xlabel('rshift (cm-1)')
ylabel('intensity (a.u.)')
title(plot_title)

% Information related to the lowest part of the spectrum.
data_L(:,1)=rshift_c_L(non_zero_elements_sample_peaks);
data_L(:,2)=sample_peaks(non_zero_elements_sample_peaks).^(1\nbOfSpectra);

```

6.2. absoluteMaxima_detection_fn.m

```
function [M_absolute_maximum] =
absoluteMaxima_detection_fn(intensity_bzn_divided_baseline_removed,indexes_vector,intensity_gscatter)

    [nbOfRows,~]=size(intensity_bzn_divided_baseline_removed);
M_absolute_maximum = zeros(nbOfRows,1);
peak_vector_absolute_maximum = 0;

    for m=numel(indexes_vector):-1:3
if ((indexes_vector(m)-indexes_vector(m-1))== 1 &&
(indexes_vector(m)-indexes_vector(m-2))== 2)
%Each peak needs to have almost three points to be considered.
peak_vector_absolute_maximum(m)=intensity_gscatter(m);
peak_vector_absolute_maximum(m-1)=intensity_gscatter(m-1);
%We keep the last two points of the peak, which are not save by the condition.
peak_vector_absolute_maximum(m-2)=intensity_gscatter(m-2);
else
    maximum_absolute_maximum,maximum_index_absolute_maximum
=
max(peak_vector_absolute_maximum);
M_absolute_maximum(maximum_index_absolute_maximum)=maximum_absolute_maximum;
peak_vector_absolute_maximum = 0;
end
end
end
```