



---

**Universidad de Valladolid**

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE  
TELECOMUNICACIÓN

GRADO EN INGENIERÍA DE TECNOLOGÍAS ESPECÍFICAS DE  
TELECOMUNICACIÓN

Mención en sistemas de Telecomunicación

**Desarrollo de algoritmia para la identificación automática  
de espectros Raman de la misión ExoMars**

Autor:

Mohamed Khadmaoui Bichouna

Tutor: Ramón de la Rosa Steinz

Cotutor: Guillermo López Reyes



(Esta página está en blanco intencionadamente)



## Comisión evaluadora

<b>Presidente</b>	J. Carlos Aguado Manzano
<b>Secretario</b>	Ramón de la Rosa Steinz
<b>Vocal</b>	Ramón J. Durán Barroso
<b>Suplente 1</b>	Alonso Alonso Alonso
<b>Suplente 2</b>	Javier M. Aguiar Pérez



(Esta página está en blanco intencionadamente)



## **Agradecimientos**

En primer lugar, me gustaría agradecer a mi tutor Guillermo Lopez-Reyes el haberme dado la oportunidad de poder participar en un proyecto tan apasionante como es la misión Exomars y trabajar en un campo de investigación como es la exploración espacial, un campo que desde pequeño me ha llamado la atención e interesado para trabajar en él y poder aportar mi granito de arena. Agradecer también sus valiosas correcciones y toda la aportación que me ha brindado.

Por otra parte, quería dar las gracias a la Unidad Asociada UVA-CSIC-CAB por dejarme desarrollar el trabajo fin de grado con ellos y haberme facilitado los medios necesarios para el desarrollo de éste.

Por último, pero no por ello menos importante, expresar gratitud a mi familia y amigos por su apoyo continuo durante todo este tiempo.

Muchas gracias a todos.



(Esta página está en blanco intencionadamente)



## RESUMEN Y OBJETIVOS

Este trabajo se ha desarrollado en el marco de desarrollo del instrumento RLS, el espectrómetro Raman incluido en el rover Rosalind Franklin que la ESA lanzará a Marte en 2022 en la misión ExoMars.

El objetivo principal del trabajo es el desarrollo de algoritmos y rutinas para el análisis automatizado de espectros Raman que permitan la identificación de las fases minerales presentes en dichos espectros. Este trabajo se enmarca en el desarrollo del software IDAT/SpectPro que, como ya se ha descrito, es una herramienta software desarrollada por la Universidad de Valladolid y que será utilizada durante la fase de operación del instrumento RLS en Marte para el análisis automatizado de los datos obtenidos de muestras marcianas. Las rutinas que se proponen facilitarán la tarea de interpretación de los espectros, cruzando los datos obtenidos contra una base de datos espectrales desarrollada ad-hoc para la misión.

Para ello, el trabajo se ha dividido en dos tareas.

- 1- Por un lado, el alumno ha de familiarizarse con el tratamiento y procesamiento de espectros Raman. En base a espectros ya adquiridos, se pretende procesar un número suficiente de espectros de materiales y mezclas de materiales diferentes que permitan realizar una caracterización adecuada de las rutinas desarrolladas. De esta forma, se obtendrán capacidades de procesamiento y análisis de espectroscopia Raman, desde el tratamiento de la línea de base, detección de picos o ajustes de bandas sobre los espectros procesados.
- 2- Por otro lado, el alumno desarrollará una serie de algoritmos de identificación en base a parámetros espectrales (posición y altura de picos) que habrán sido previamente obtenidos mediante el análisis con SpectPro definido en el objetivo 1. Finalmente, se testeará y parametrizará el algoritmo desarrollado utilizando los espectros procesados de minerales puros y mezclas obtenidos en el laboratorio. Esto permitirá obtener una rutina de identificación optimizada para un conjunto heterogéneo de espectros.



(Esta página está en blanco intencionadamente)



## ABSTRACT AND OBJECTIVES

This work has been developed within the framework of the development of the RLS instrument, the Raman spectrometer included in the Rosalind Franklin rover that ESA will launch to Mars in 2022 on the ExoMars mission.

The main objective of the work is the development of algorithms and routines for the automated analysis of Raman spectra that allow the identification of the mineral phases present in these spectra. This work is part of the development of the IDAT/SpectPro software which is a software tool developed by the University of Valladolid that will be used during the operation phase of the RLS instrument on Mars for on-ground automated analysis of data obtained from Martian samples. The proposed routines will facilitate the task of interpreting the spectra, crossing the data obtained against a spectral database developed ad-hoc for the mission.

To do this, the work has been divided into two tasks:

1. On one hand, the student must become familiar with the treatment and processing of Raman spectra. Based on already acquired spectra, it is intended to process a sufficient number of spectra of materials and mixtures of different materials that allow to perform an adequate characterization of the developed routines. In this way, the student will gain capabilities on the processing and analysis of Raman spectroscopy data, from baseline treatment, peak detection or band adjustments on processed spectra.
2. On the other hand, the student will develop a series of identification algorithms based on spectral parameters (position and height of peaks) that will help identify the material on a determined spectrum by comparing with the spectral database, providing the best-estimate results. Finally, the developed algorithm will be tested and parameterized by using the processed spectra of pure minerals and mixtures obtained in the laboratory. This will allow for an optimized identification routine for a heterogeneous set of spectra.



(Esta página está en blanco intencionadamente)



## Índice de contenido

1. Introducción .....	18
1.1 La búsqueda de vida en Marte.....	18
1.2 La misión ExoMars.....	19
1.2.1 Operando un rover en la superficie de Marte .....	19
1.2.2 La instrumentación del rover ExoMars 2020 .....	20
1.3 La espectroscopia Raman y el instrumento RLS.....	22
1.4 Operación y procesamiento de datos de RLS. IDAT y SpectPro .....	24
2 Materiales y métodos .....	26
2.1 Base de datos de espectros.....	26
2.2 Selección de los espectros problema .....	30
2.3 Pre-tratamiento de los espectros problema .....	30
2.3.1 Tratamiento espectral.....	30
2.3.2 Creación de los ficheros de identificación.....	34
2.4 Propuesta de algoritmo de identificación automática.....	34
2.5 Filosofía de pruebas y parametrización del algoritmo.....	37
3 Resultados y discusión .....	38
3.1 Espectros puros.....	39
3.2 Espectros mezcla.....	41
3.3 Combinación de todos los espectros .....	43
4 Conclusiones y líneas futuras .....	44
5 Anexos.....	46
5.1 Anexo 1: Descripción del algoritmo .....	46
5.1.1 Desarrollo de rutina de identificación.....	46
5.1.2 Matlab .....	46
5.1.3 Lectura de la base de datos.....	47
5.1.4 Identificación de los materiales .....	48
5.1.5 Lectura del espectro problema .....	48
5.2 Anexo 2 - Verificación del script.....	52
5.3 Anexo 3: Gráficas de espectros puros.....	56
5.4 Anexo 4: Gráficas de espectros mezcla.....	61
5.5 Anexo 5: Gráficas utilizando todos los espectros problema.....	67
5.6 Anexo 6: Análisis detallado de algunos espectros individuales .....	72



5.6.1	Calcita .....	72
5.6.2	Olivino .....	74
5.6.3	Cuarzo y Calcita .....	76
6	Bibliografía .....	81



(Esta página está en blanco intencionadamente)



## Índice de figuras

Figura 1. El rover Rosalind Franklin de la misión ExoMars 2022 y sus instrumentos .....	22
Figura 2. Espectro Raman de una mezcla de calcita y yeso .....	23
Figura 3. Instrumento RLS .....	24
Figura 4. Interfaz de SpectPro .....	25
Figura 5. Espectro de la calcita .....	27
Figura 6. Espectro de yeso .....	28
Figura 7. Espectro de una mezcla de calcita y yeso, donde pueden observarse picos de ambos materiales.....	28
Figura 8. Espectro recortado .....	32
Figura 9. Parámetros para la eliminación de línea de base .....	32
Figura 10. Bandas detectadas por el algoritmo en un espectro de calcita y cuarzo.....	33
Figura 11. Ejemplo de archivo xml .....	34
Figura 12.[P-B-I] = [0.3-0.3-0.4] .....	41
Figura 13.[P-B-I] = [0.1-0.3-0.6] .....	41
Figura 14.[P-B-I] = [0.1-0.5-0.4] .....	42
Figura 15.[P-B-I] = [0.1-0.3-0.6] .....	42
Figura 16.[P-B-I] = [0.1-0.3-0.6] .....	44
Figura 17. Script A .....	47
Figura 18. Script para obtener datos de un archivo xml .....	48
Figura 19.Lectura del espectro problema .....	48
Figura 20. Creación de la lista con posibles materiales.....	49
Figura 21.Código de la función para la creación de la lista de posibles materiales.....	49
Figura 22.Orden de la lista .....	50
Figura 23.Código de la función de coincidencias .....	51
Figura 24.Eliminación de bandas.....	52
Figura 25.Lectura de espectros de la base de datos .....	53
Figura 26.Lectura de archivo xml y archivo txt .....	53
Figura 27.Valor de parámetros .....	54
Figura 28.Bucle para probar los distintos valores de parámetros .....	55
Figura 29.Indicadores generales .....	55
Figura 30.Código para mostrar gráficas .....	56
Figura 31. Combinación [P-B-I] = [0.1-0.9-0] .....	56
Figura 32. [P-B-I] = [0.3-0-7-0] .....	57
Figura 33.[P-B-I] = [0.5-0.5-0] .....	57
Figura 34.[P-B-I] = [0.7-0.3-0] .....	57
Figura 35.[P-B-I] = [0.9-0.1-0] .....	58
Figura 36. [P-B-I] = [0.1-0.7-0.2] .....	58
Figura 37.[P-B-I] = [0.3-0.5-0.2] .....	58
Figura 38.[P-B-I] = [0.5-0.3-0.2] .....	59
Figura 39.[P-B-I] = [0.7-0.1-0.2] .....	59
Figura 40.[P-B-I] = [0.1-0.5-0.4] .....	59
Figura 41.[P-B-I] = [0.3-0.3-0.4] .....	60
Figura 42.[P-B-I] = [0.5-0.1-0.4] .....	60
Figura 43.[P-B-I] = [0.1-0.3-0.6] .....	60



Figura 44.[P-B-I] = [0.3-0.1-0.6] .....	61
Figura 45.[P-B-I] = [0.1-0.1-0.8] .....	61
Figura 46.[P-B-I] = [0.1-0.9-0] .....	62
Figura 47.[P-B-I] = [0.3-0.7-0] .....	62
Figura 48.[P-B-I] = [0.5-0.5-0] .....	62
Figura 49.[P-B-I] = [0.7-0.3-0] .....	63
Figura 50.[P-B-I] = [0.9-0.1-0] .....	63
Figura 51.[P-B-I] = [0.1-0.7-0.2] .....	63
Figura 52.[P-B-I] = [0.3-0.5-0.2] .....	64
Figura 53.[P-B-I] = [0.5-0.3-0.2] .....	64
Figura 54.[P-B-I] = [0.7-0.1-0.2] .....	64
Figura 55.[P-B-I] = [0.1-0.5-0.4] .....	65
Figura 56.[P-B-I] = [0.3-0.3-0.4] .....	65
Figura 57.[P-B-I] = [0.5-0.1-0.4] .....	65
Figura 58.[P-B-I] = [0.1-0.3-0.6] .....	66
Figura 59.[P-B-I] = [0.3-0.1-0.6] .....	66
Figura 60.[P-B-I] = [0.1-0.1-0.8] .....	66
Figura 61. [P-B-I] = [0.1-0.9-0] .....	67
Figura 62.[P-B-I] = [0.3-0.7-0] .....	67
Figura 63.[P-B-I] = [0.5-0.5-0] .....	67
Figura 64.[P-B-I] = [0.7-0.3-0] .....	68
Figura 65.[P-B-I] = [0.9-0.1-0] .....	68
Figura 66.[P-B-I] = [0.1-0.7-0.2] .....	68
Figura 67.[P-B-I] = [0.3-0.5-0.2] .....	69
Figura 68.[P-B-I] = [0.5-0.3-0.2] .....	69
Figura 69.[P-B-I] = [0.7-0.1-0.2] .....	69
Figura 70.[P-B-I] = [0.1-0.5-0.4] .....	70
Figura 71.[P-B-I] = [0.3-0.3-0.4] .....	70
Figura 72.[P-B-I] = [0.5-0.1-0.4] .....	70
Figura 73.[P-B-I] = [0.1-0.3-0.6] .....	71
Figura 74.[P-B-I] = [0.3-0.1-0.6] .....	71
Figura 75.[P-B-I] = [0.1-0.1-0.8] .....	71
Figura 76. Espectro base de datos y Espectro problema del material olivino .....	76
Figura 77. Espectro de la base de datos de la calcita .....	79
Figura 78. Espectro de la base de datos del cuarzo .....	79
Figura 79. Archivo xml del cuarzo .....	80
Figura 80. Espectro de la mezcla calcita y cuarzo .....	80



(Esta página está en blanco intencionadamente)



## Índice de tablas

Tabla 1. Materiales de la base de datos.....	27
Tabla 2. Espectros problema provenientes de materiales puros.....	29
Tabla 3. Espectros problema provenientes de mezclas de 2 materiales .....	30
Tabla 4. Espectros problema provenientes de mezclas de 3 materiales .....	30
Tabla 5. Resultados de la comparación de bandas .....	75
Tabla 6. Resultados de la comparación de las bandas.....	77
Tabla 7. Resultados de la segunda interacción .....	78



(Esta página está en blanco intencionadamente)



## 1. Introducción

### 1.1 La búsqueda de vida en Marte

La búsqueda de vida en otros planetas es vital para conocer nuestro origen y esta búsqueda empieza en Marte. Si alguna vez ha existido vida en él tuvo que ser cuando éste era más cálido y húmedo, cuando tenía condiciones similares a las de la tierra hace unos 4000 millones de años.

Uno de los problemas del planeta Tierra es que es muy difícil encontrar rocas que tengan más de 3 mil millones de años en buenas condiciones debido a los procesos metamórficos de altas temperaturas y la tectónica de placas. Pero Marte no ha sufrido esta actividad debido a su menor tamaño y, por lo tanto, es posible encontrar muestras de formaciones rocosas de un periodo más antiguo que puedan ser accesibles para analizar.

La vida microbiana se extendió muy rápidamente sobre la superficie de la Tierra cuando apareció. Considerando la similitud de las condiciones de Marte y la Tierra primigenias, algo así podría haber ocurrido en Marte. La misión ExoMars a Marte tiene como objetivo la identificación de esta posible vida, que supondría uno de los descubrimientos más trascendentales del ser humano. No obstante, para esto es también necesario verificar que existen o existieron lugares favorables a la vida con agua líquida al menos durante ciertos periodos de la historia geológica marciana. Desgraciadamente, en las condiciones actuales del planeta es muy improbable encontrar vida presente, por lo que la búsqueda se centra en la detección de vida extinta.

Para ello, las misiones robóticas a Marte buscan dos tipos de estructuras relacionadas con la vida: morfológicas y químicas.

En la superficie de las rocas podemos encontrar información morfológica relacionada con los procesos biológicos como, por ejemplo: capas bacterianas fosilizadas, estromatolitos, etc. Para encontrar estas evidencias hace falta un sistema de imágenes capaz de cubrir un rango de distancias bastante amplio.

Para tener una identificación química eficaz de biomarcadores se necesita acceder a moléculas orgánicas bien conservadas. Esto supone un problema ya que las condiciones de la superficie de Marte suponen una serie de efectos adversos, relacionados entre otras cosas con la atmósfera tan tenue y las altas dosis de radiación que tiene Marte, por lo que la preservación de estos biomarcadores y moléculas orgánicas complejas se ve gravemente afectada en la superficie y en los primeros centímetros bajo la superficie. La dosis de rayos ultravioleta es más alta que en el planeta Tierra, así como la radiación ionizante y no ionizante, que, en ausencia de un campo magnético, son intensas y dañan fácilmente a los organismos o biomoléculas que potencialmente pudieran encontrarse en su superficie.



La fotoquímica inducida por rayos ultravioleta es responsable de la producción de especies oxidantes reactivas que pueden destruir biomarcadores si se activan. Además, la radiación ionizante penetra en los metros más altos del subsuelo del planeta que provoca un proceso de degradación, que tras millones de años de exposición puede alterar las moléculas orgánicas más allá de la sensibilidad de detección de los instrumentos analíticos. Los materiales más cercanos a la superficie son los que están expuestos a una dosis más alta.

## **1.2 La misión ExoMars**

ExoMars es un programa conjunto de la Agencia Espacial Europea (ESA) y la agencia rusa Roscosmos cuyo objetivo es investigar el entorno de Marte. Los objetivos del programa son: la búsqueda de signos de vida presente o pasada en Marte; investigación de las variaciones del agua y el ambiente geoquímico y, por último, la investigación de los gases de la atmósfera y su procedencia. El programa ExoMars está compuesto por 2 misiones.

La primera de ellas, llamada Trace Gas Orbiter (TGO), fue lanzada en 2016 transportando un satélite, que orbita a una distancia de 400 km aproximadamente de la superficie de Marte, y un módulo de entrada y aterrizaje (Schiaparelli). El objetivo de esta primera misión es detectar y estudiar los gases que componen la atmósfera de Marte, sobre todo, el gas metano. Este gas es importante ya que podría revelar la actividad de procesos biológicos o geológicos, ya que este elemento tiene una duración de vida pequeña y su presencia sugiere la existencia de procesos que podrían seguir activos actualmente. El módulo de aterrizaje constaba de varios sensores e instrumentos para estudiar el entorno del lugar de aterrizaje, aunque lamentablemente falló en el procedimiento de aterrizaje y éste se estrelló de forma catastrófica contra la superficie de Marte [1].

La segunda misión del programa ExoMars se compone de un vehículo de exploración (rover) bautizado Rosalind Franklin (ver Figura 1), cuyo objetivo principal será la búsqueda de vida. Para ello, el rover porta un taladro para obtener muestras a dos metros de profundidad, donde se espera que las muestras orgánicas puedan haberse preservado al estar más protegidas de la radiación de Marte. Además, el rover incorpora una serie de instrumentos para realizar investigaciones astrobiológicas y geoquímicas. El lanzamiento de esta misión está previsto en 2022 [2].

### **1.2.1 Operando un rover en la superficie de Marte**

Para que la misión sea exitosa hay que encontrar el lugar adecuado para situar el rover y que éste consiga aterrizar de una forma segura. El lugar se puede identificar gracias a la información de teledetección de los orbitadores que ya se encuentran estudiando Marte. Dado que la misión de ExoMars es la búsqueda de vida, los posibles sitios para el aterrizaje han de contener evidencias que sugieran un entorno habitable o que haya



sido habitable en algún momento en el pasado por información morfológica y mineralógica [3].

La estrategia que se va a seguir para lograr los objetivos es:

- Aterrizar en una ubicación que posea un alto interés exobiológico por signos de vida pasadas o presentes.
- Recolectar muestras científicas de diferentes sitios con un taladro capaz de llegar al subsuelo y afloramientos rocosos de la superficie.
- Llevar una serie de mediciones como son: evaluación del entorno geológico, investigaciones en afloramientos superficiales y colección de subsuelos bien seleccionados para ser estudiados en el laboratorio analítico del Rover.

Un detalle a tener en cuenta es si, por algún casual, se detectan compuestos orgánicos en Marte, hay que demostrar que estos compuestos no han sido llevados de la Tierra. Para garantizarlo, el rover llevará consigo una serie de muestras de calibración en blanco que analizará después de aterrizar.

Una vez ahí, una de las primeras acciones será pasar una muestra en blanco al laboratorio analítico y proceder a su análisis. Siendo muestras estériles y sin presencia de orgánicos, los resultados deben ser coherentes y no mostrar trazas de vida o materia orgánica de cualquier tipo. Si los resultados obtenidos distan de lo que se espera, cualquier hallazgo de búsqueda vida que se pueda encontrar podría quedar invalidado.

Durante los 180 soles (días marcianos) que durará la misión, el rover ExoMars realizará 6 ciclos de experimentos que conllevarán la adquisición y análisis de una muestra de superficie y una muestra del subsuelo. Además, realizará durante el tiempo de misión dos muestreos verticales consistentes en obtener y analizar 5 muestras del subsuelo en incrementos de 50 cm de profundidad, de 0 cm a 200 cm, para poder estudiar las variaciones con la profundidad.

### 1.2.2 La instrumentación del rover ExoMars 2020

Para cumplir sus objetivos, el rover Rosalind Franklin (Figura 1), llevará consigo los siguientes instrumentos [4]:

- PanCam (The **P**anoramic **C**amera): una cámara panorámica con la que se obtendrá un mapa del terreno digital.
- ISEM (**I**nfrared **S**pectrometer for **E**xo**M**ars): con este instrumento se podrá evaluar la composición mineralógica en superficie. Al trabajar con PanCam, ISEM contribuirá a la selección de muestras adecuadas para su posterior análisis por los otros instrumentos.



- CLUPI (**C**lose – **U**P **I**mager): es un sistema de cámara para adquirir imágenes de primer plano en color de alta resolución de rocas, afloramientos, finos de perforación y las muestras obtenidas con el taladro.
- WISDOM (**W**ater **I**ce and **S**ubsurface **D**eposit **O**bservation **O**n **M**ars): Un radar de penetración terrestre para caracterizar la estratigrafía bajo el rover. WISDOM se utilizará con Adron, que puede proporcionar información sobre el contenido de agua del subsuelo, para decidir dónde recolectar muestras del subsuelo para su análisis.
- Adron: instrumento para la búsqueda de agua subterránea y minerales hidratados. Adron se utilizará con WISDOM.
- Ma\_MISS (**M**ars **M**ultispectral **I**mager for **S**ubsurface **S**tudies): se ubica en la punta del taladro, y contribuirá al estudio de la mineralogía y formación rocosa de Marte en las zonas donde se extraerán las muestras.
- MicrOmega: Uno de los tres instrumentos del laboratorio analítico. Es un espectrómetro de imágenes visible e infrarrojo para estudios de mineralogía en las muestras obtenidas por el taladro.
- MOMA (**M**ars **O**rganic **M**olecule **A**nalyser): Otro de los instrumentos del laboratorio analítico, se centrará en el estudio de biomarcadores y orgánicos con bajos límites de detección para responder preguntas relacionadas con el origen, la evolución y la distribución potenciales de la vida en Marte.
- RLS (Raman spectrometer): el tercer instrumento del laboratorio analítico que permitirá caracterizar la composición mineralógica e identificar pigmentos orgánicos en las muestras obtenidas por el taladro.

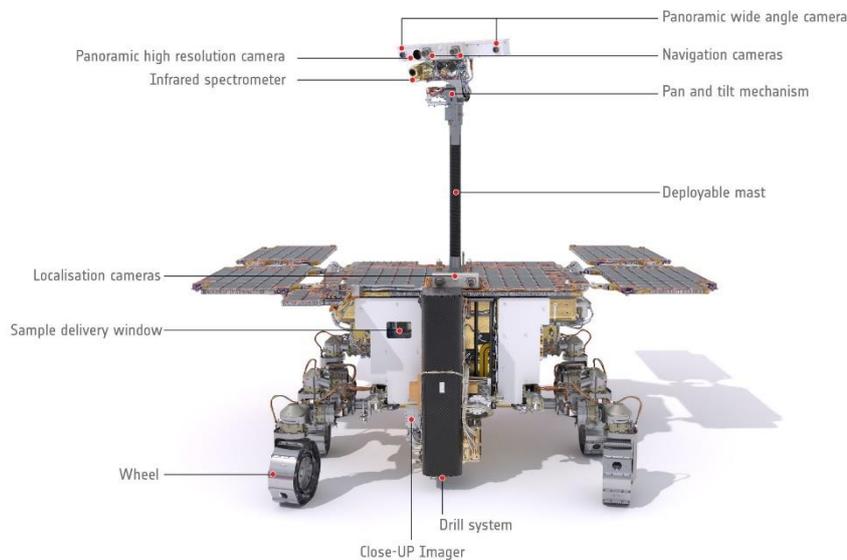


Figura 1. El rover Rosalind Franklin de la misión ExoMars 2022 y sus instrumentos

### 1.3 La espectroscopia Raman y el instrumento RLS

La espectrometría Raman es una técnica espectroscópica que proporciona, en poco tiempo y de forma no destructiva, información química y estructural (molecular) de cualquier compuesto orgánico o inorgánico permitiendo su identificación [5]. La técnica consiste en el análisis de la luz dispersada por un material al ser éste excitado mediante un haz de luz monocromático. La mayor parte de la luz dispersada tiene la misma frecuencia que la luz incidente pero una fracción de ésta presenta un cambio de frecuencia resultado de la interacción del material con la luz. Esta pequeña fracción es la que proporciona información sobre la composición del material analizado y es conocida como “dispersión Raman” [5].

Por sus características, la espectroscopia Raman tiene un gran potencial analítico que, junto con desarrollos tecnológicos que han permitido la miniaturización y la mejora de los instrumentos, se ha traducido en su uso en varios campos de investigación o industriales, incluyendo en la actualidad misiones planetarias tales como Mars 2020 de la NASA, lanzada en julio de 2020, donde hay dos espectrómetros Raman, (uno de ellos, SuperCam, donde el grupo de investigación con el que se realiza este trabajo participa en el grupo de ciencia y en el desarrollo del sistema de calibración), o el rover Rosalind Franklin que lanzará la ESA a Marte en 2022 e incorpora el instrumento RLS [6].

El espectro Raman (figura 2) representa las diferentes frecuencias dispersadas por el material en función del número de onda normalizado con respecto al número de onda de excitación, en una magnitud referida como “desplazamiento Raman”, y medida en  $\text{cm}^{-1}$ .

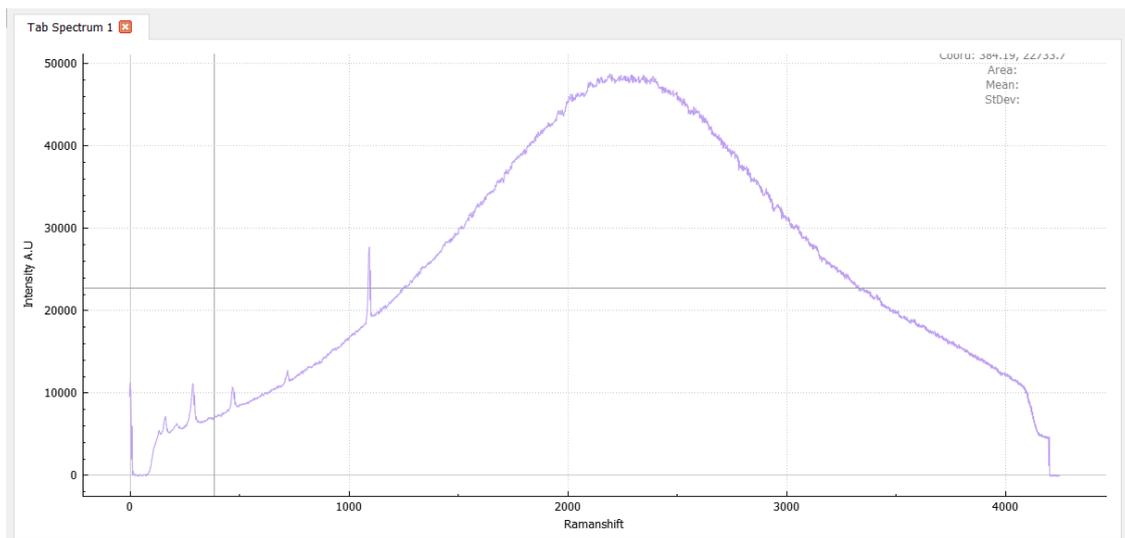


Figura 2. Espectro Raman de una mezcla de calcita y yeso

El instrumento RLS o Raman Laser Spectrometer (figura 3) [7] es un espectrómetro Raman, constituyendo una poderosa herramienta para la identificación y la caracterización de minerales y biomarcadores. Como ya se ha mencionado, la espectroscopia Raman es sensible a la composición y estructura de cualquier compuesto mineral u orgánico, de manera rápida y no destructiva. Esta capacidad se utilizará para detectar e identificar compuestos orgánicos potenciales que puedan estar relacionados con trazas de vida presente o pasadas, así como para extraer información mineralógica para poder caracterizar procesos ígneos, metamórficos y sedimentarios.

El instrumento RLS se utilizará para [8]:

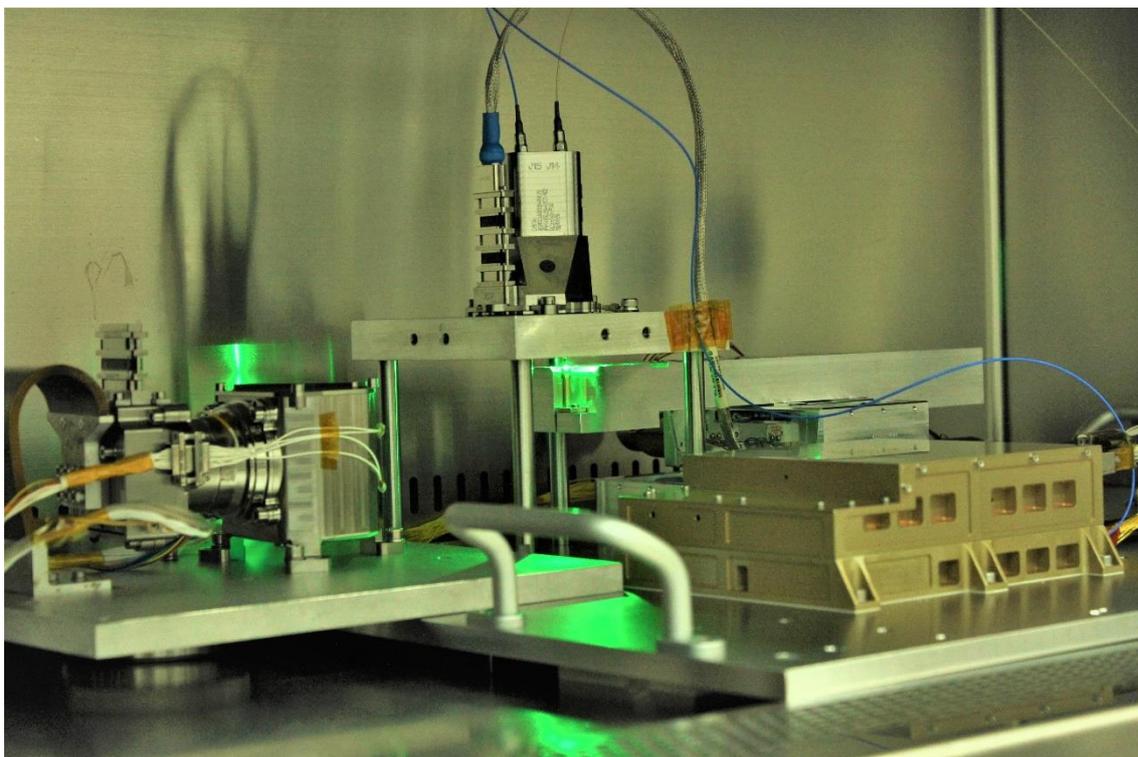
- Identificar compuestos orgánicos y buscar signos de vida.
- Identificar minerales e indicadores de actividades biológicas.
- Caracterizar las fases de los minerales producidas por procesos relacionados con el agua.
- Caracterizar minerales ígneos y sus productos resultantes de procesos de alteración.

RLS, formando parte del laboratorio analítico del rover, funcionará de forma cooperativa con MicrOmega y MOMA, los otros instrumentos del laboratorio analítico del rover, para el análisis de las muestras extraídas de la subsuperficie de Marte. De esta forma la información obtenida por el instrumento RLS se correlacionará con la obtenida por los otros instrumentos, en un escenario sin precedentes, ya que el diseño implementado en el rover permite a estos tres instrumentos realizar análisis en el mismo punto de la muestra, en lo que se conoce como Ciencia Colaborativa [8].

La espectroscopia Raman consiste en analizar la luz dispersa emitida por una muestra iluminada por un láser que induce vibraciones propias de los enlaces moleculares de la



muestra. La información espectral que recibe el espectrómetro está relacionada con la estructura molecular y la composición de un compuesto, lo que permite la identificación y caracterización de las moléculas, minerales y rocas marcianas. Como ya se ha mencionado, la espectroscopia Raman permite analizar la muestra de forma no destructiva, lo que significa que la muestra permanece intacta para su posterior análisis.



*Figura 3. Instrumento RLS*

#### **1.4 Operación y procesamiento de datos de RLS. IDAT y SpectPro**

Durante la fase de operación de la misión los datos de misión serán recibidos en el Centro de Operaciones del rover (ROCC), situado en Turín (Italia). Desde que los datos provenientes de Marte llegan al ROCC, se dispone únicamente de unas pocas horas para analizarlos, interpretarlos y definir el plan de operación del siguiente sol. Por tanto, la toma de decisiones tácticas debe ocurrir de la forma más rápida posible.

En este sentido, el equipo del instrumento RLS está desarrollando un software de recepción y tratamiento de datos de ingeniería y ciencia del instrumento, denominado IDAT (Instrument Data Analysis Tool -para datos de ingeniería-) y SpectPro (en su desarrollo para productos científicos).

IDAT es una herramienta de software que se utiliza para la recepción, decodificación, calibración y verificación de las telemetrías generadas por el instrumento RLS. Con esto



se puede verificar el estado del instrumento, la recepción e interpretación de los datos de una manera fácil y rápida [9].

Por otro lado, SpectPro (figura 4) es una segunda instancia de IDAT que permite realizar el análisis de datos científicos (espectros) para su interpretación y análisis gracias a la incorporación de múltiples herramientas de operación sobre espectros que permiten a los científicos procesar los espectros y así permitirá obtener análisis detallados en base a los datos recibidos del instrumento [9].

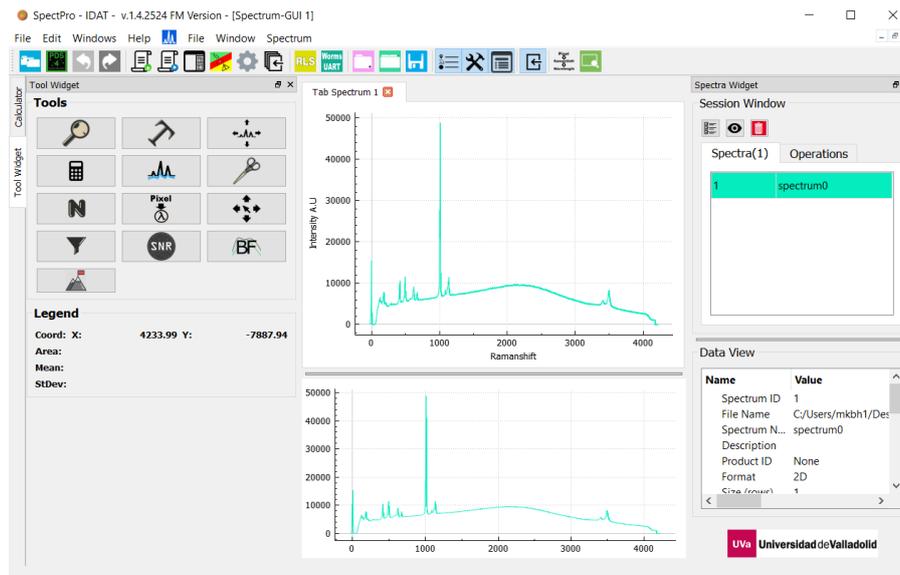


Figura 4. Interfaz de SpectPro

La utilización de IDAT/SpectPro durante los ensayos e integración del instrumento RLS han demostrado su utilidad permitiendo la interpretación “en vivo” de los datos y la monitorización del estado del instrumento durante toda la fase de desarrollo. Por otro lado, las herramientas analíticas para los espectros adquiridos (SpectPro) han facilitado un rápido y eficaz análisis de los datos obtenidos durante la caracterización del instrumento.

De cara a la operación en Marte, IDAT/SpectPro continúa su desarrollo para adaptar las interfaces al funcionamiento con los pipelines de datos del ROCC, pero también para desarrollar nuevas funciones de análisis, así como rutinas automatizadas que permitan obtener la mayor cantidad de información posible a partir de los espectros. El objetivo es que IDAT/SpectPro genere informes automáticos con una interpretación preliminar. De esta forma se reduce el tiempo dedicado a procesamiento durante los ciclos tácticos de operación, donde la toma de decisiones tendrá que ser lo más rápida posible.

Es en este ámbito de desarrollo automatizado donde se enmarca el desarrollo de este Trabajo de Fin de Grado. La identificación de las fases minerales que conforman un espectro es la base sobre la que los científicos pueden tomar decisiones interpretativas de la geología marciana, y de esa manera tomar decisiones en cuanto a los planes de operación del instrumento y el rover. Por tanto, se plantea un trabajo encaminado al



desarrollo de rutinas de análisis que permitan realizar la interpretación espectral en base a espectros Raman. Es decir, el desarrollo de algoritmia para realizar la identificación de los materiales que conforman un espectro a partir de la información de las bandas de dichos espectros. Para ello, las rutinas compararán los espectros obtenidos con el instrumento con respecto a una base de datos de referencia de muestras potencialmente representativas de la geología marciana, evaluando la capacidad de detección e identificación de dichas rutinas.

## 2 Materiales y métodos

### 2.1 Base de datos de espectros

Para el desarrollo de la rutina de identificación de materiales se ha contado con una serie de espectros ya adquiridos por el sistema RLS u otros espectrómetros Raman utilizados en el marco de desarrollo del instrumento RLS. Dichos espectros están divididos en dos grupos: por un lado, una base de datos de estándares geológicos, y por otro un conjunto de “espectros problema” provenientes de muestras reales, que se utilizarán para desarrollar y verificar el funcionamiento de los algoritmos de identificación desarrollados.

La base de datos de materiales forma parte de un proyecto abierto del grupo de investigación para el desarrollo de ésta, por lo que no contiene todos los posibles materiales. Actualmente la base de datos cuenta con una población de 53 muestras, todas ellas procesadas adecuadamente (tal y como se describe en las siguientes secciones), estando disponible toda la información espectral necesaria para la ejecución de las rutinas.

La Tabla 1 muestra todos los materiales presentes actualmente en la base de datos.

Actinolita	Hematita
Alunita	Hidrocerusita
Anatasa	Jarosita
Anglesita	Caolinita
Anhidrita	Lepidolita
Antigorita	Magnesita
Apatita	Magnetita
Aragonita	Manganeso
Barita	MgSulfate
Berilo	NaSulfate
Calcita	Ortoclasa
Carbon	Pirita
Grafito	Pirolusita
Celestina	Piroxeno
Cromita	Cuarzo



Crisotila	Rodocrosita
Copiapita	Rozenita
Crocoita	Rutila
Dolomita	Serpentina
Epidota	Siderita
Fluorita	Silicio
Forstorita	Sodalita
Galena	Esfalerita
Goethita	Azufre
Grosular	Talco
Yeso	Vermiculita
Halita	

Tabla 1. Materiales de la base de datos

El segundo conjunto de espectros son los espectros problema que se han utilizado para para probar y verificar el código y las rutinas de identificación automática. Este grupo de espectros se divide en 3 grupos según el número de materiales presentes en el espectro: puros (espectros correspondientes a materiales en solitario, como los ejemplos mostrados en las figuras 5 y 6, de calcita y yeso, respectivamente), mezclas binarias (espectros resultantes de muestras con 2 materiales donde se pueden observar las características espectrales de ambos materiales - ejemplo en figura 7 para una mezcla de calcita y yeso) y mezclas ternarias (3 materiales presentes simultáneamente en el espectro).

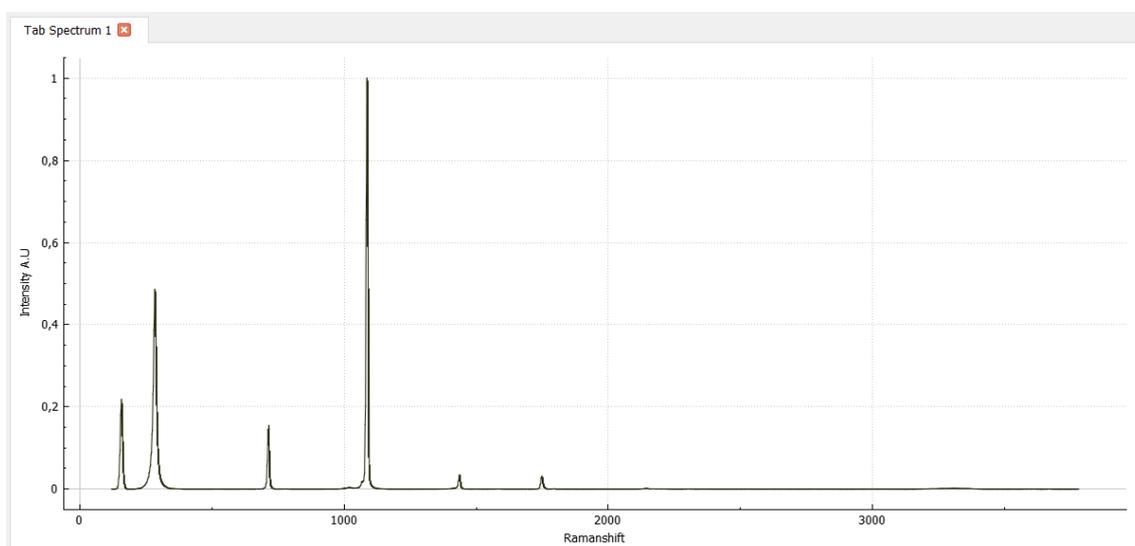


Figura 5. Espectro de la calcita

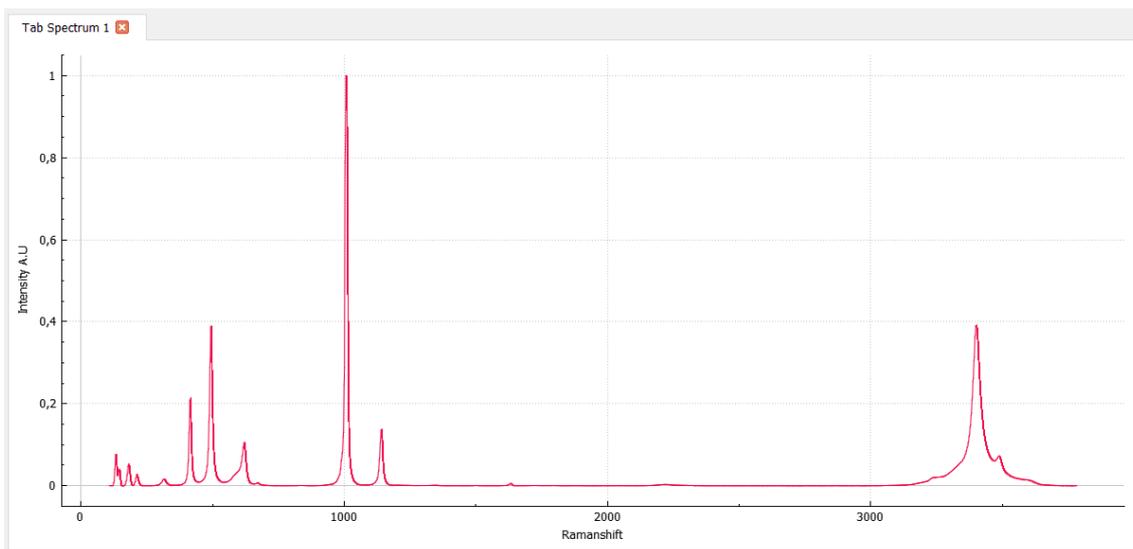


Figura 6. Espectro de yeso

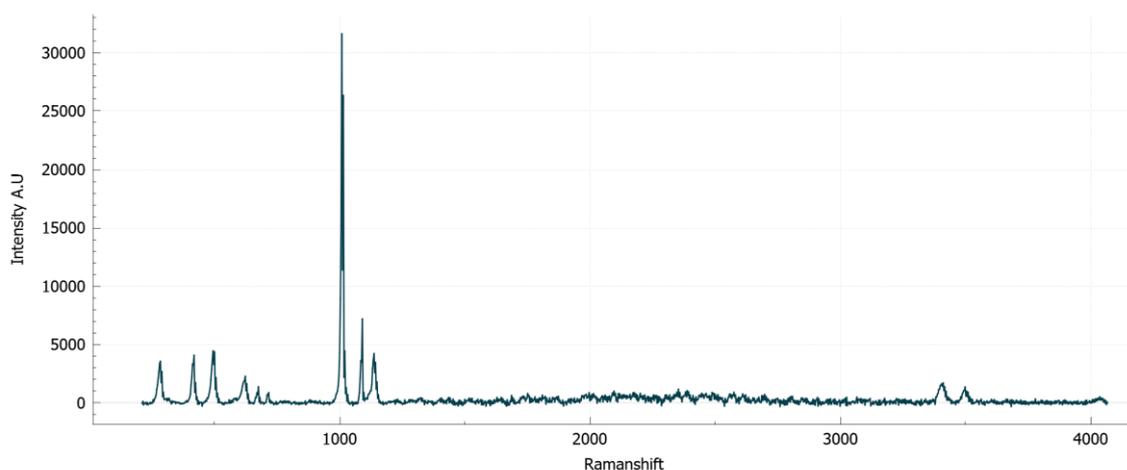


Figura 7. Espectro de una mezcla de calcita y yeso, donde pueden observarse picos de ambos materiales.

La lógica detrás de la elección de estos materiales es verificar la capacidad de los algoritmos de identificar no solamente espectros de muestras puras, sino también la capacidad de identificación de las múltiples fases minerales potencialmente presentes en un espectro (lo cual es habitual en los análisis de muestras heterogéneas como los que realizará el instrumento RLS).

Los espectros puros están clasificados en 7 grupos diferentes según se recoge en la Tabla 2. Los espectros de mezclas binarias están indicados en la tabla 3. Finalmente, los espectros de mezclas ternarias están formados por olivinos, piroxeno y feldespatos con concentraciones variables, tal y como se recoge en la tabla 4.

Carbonatos	Calcita
------------	---------



Feldespatos	Albita
	Amazonita
	Andesina
	Anortita
	Anorthoclase
	Celsiana
	Labradorita
	Microclina
	Oligoclase
	Ortoclase
	Sanidina
Olivino	Olivino
Piroxeno	Piroxeno
Cuarzo	Cuarzo
Serpentina	Serpentina
Sulfatos	Anhidrita
	CaSulfate
	Yeso
	MgSulfate
	NaSulfate

Tabla 2. Espectros problema provenientes de materiales puros

Mezcla	Concentraciones
Anhidrita_NaSulfate	Anhidrita25_NaSulfate75
	Anhidrita50_NaSulfate50
	Anhidrita75_NaSulfate25
Calcita_Yeso	Calcita25_Yeso75
	Calcita50_Yeso50
	Calcita75_Yeso25
Calcita_Cuarzo	Calcita25_Cuarzo75
	Calcita50_Cuarzo50
	Calcita75_Cuarzo25
Feldespatos_Olivino	Feldespatos25_Olivino75
	Feldespatos50_Olivino50
	Feldespatos75_Olivino25
Feldespatos_Piroxeno	Feldespatos25_Piroxeno75
	Feldespatos50_Piroxeno50
	Feldespatos75_Piroxeno25
Melanterita_NaSulfate	Melanterita50_NaSulfate50
Melanterita_Szomolnokita	Melanterita50_Szomolnokita50
Melanterita_Yeso	Melanterita50_Yeso50
MgSulfate_NaSulfate	MgSulfate50_NaSulfate50



Olivino_Piroxeno	Olivino25_Piroxeno75
	Olivino50_Piroxeno50
	Olivino75_Piroxeno25
Rozenita_Melanterita	Rozenita50_Melanterita50
Rozenita_Szomolnokita	Rozenita50_Szomolnokita50
Serpentina_Olivino	Serpentina50_Olivino50
	Serpentine75_Olivino25

Tabla 3. Espectros problema provenientes de mezclas de 2 materiales

Muestra	mg Oli	mg Pir	mg Feld	total	% Oli	% Pir	% Feld	% tot
MPC 1	167	156	118	441	37,87	35,37	26,76	100
MPC 2	186	42	202	430	43,26	9,77	46,98	100
MPC 3	96	196	288	580	16,55	33,79	49,66	100
MPC 4	209	288	73	570	36,67	50,53	12,81	100
MPC 5	195	66	297	558	34,95	11,83	53,23	100

Tabla 4. Espectros problema provenientes de mezclas de 3 materiales

## 2.2 Selección de los espectros problema

A pesar de contar con todos los espectros detallados anteriormente, para comprobar el funcionamiento de la rutina no se han podido utilizar todos los espectros problema debido a que algunos materiales de los que componen los espectros de doble o triple mezcla e incluso alguno puros, no se encuentran en la base de datos. Como se ha indicado, siendo una base de datos incompleta por encontrarse en desarrollo, esta situación era esperable. Por tanto, antes de la realización de las pruebas se hizo una preselección de los espectros problema (si bien la mayoría de los espectros fueron procesados para extraer su información espectral utilizando la metodología recogida en las siguientes secciones).

Los materiales ausentes de la base de datos son: rozenita, melanterita, szomolnokita y serpentina. Además, de entre todos los minerales del grupo de los feldespatos, en la base de datos de estándares solo se encontraba la ortoclasa, por lo que se ha debido realizar una selección cuidadosa en todos aquellos espectros problema conteniendo feldespato, ya que podía dar problemas en el resultado debido no al funcionamiento del algoritmo, sino a las diferencias espectrales entre distintos tipos de feldespatos.

## 2.3 Pre-tratamiento de los espectros problema

### 2.3.1 Tratamiento espectral



Al contrario que los espectros de la base de datos de estándares, los espectros problema utilizados en este trabajo no estaban suficientemente procesados, y su información espectral de bandas (intensidad y posición de los picos) no estaba convenientemente extraída. Por tanto, de todos los espectros problema ha sido necesario realizar un procesamiento espectral o pre-tratamiento resultante en la creación de ficheros xml que contienen la información de bandas, así como de ficheros txt con la información de los materiales presentes en el espectro.

Para realizar el análisis de la información de bandas de los espectros se ha utilizado el software SpectPro creado por la unidad UVA-CSIC-CAB. Entre otras herramientas, SpectrPro incluye un calculador de SNR, operación de eliminación de la línea de base, eliminación de rayos cósmicos, filtrado, recorte, detección de picos y ajuste de bandas a funciones gaussianas y/o lorentzianas.

Los datos relevantes de los espectros, necesarios para las rutinas de identificación creadas en este trabajo, es la información de las bandas que conforman un espectro. Es decir, la información de la posición e intensidad de los picos presentes en un espectro. Por tanto, para realizar el correcto procesamiento de los espectros problema (que estaban en bruto), las herramientas de SpectPro que han sido utilizadas han sido las de: recorte, eliminación de la línea de base, normalización del espectro, detección de picos y, por último, ajuste de los picos a bandas y exportación de la información a un fichero xml (formato bds). Este proceso se detalla en los siguientes puntos.

#### **2.3.1.1 Verificación de la calibración**

Al tener un espectro problema como el de la figura 2, lo primero que se debe hacer es observar que esté bien calibrado. Esto se hace observando que en el valor 0 del eje de abscisas se encuentre el primer pico (correspondiente a la longitud de onda de excitación, que siempre ha de ser cero ya que los números de onda de desplazamiento Raman siempre están referidos a la longitud de onda de excitación). En el caso del ejemplo, se aprecia como el espectro se encuentra perfectamente calibrado.

Si esto no fuera así habría que recalibrar el espectro de forma que la longitud de onda de excitación coincidiera con la teórica (cero). De no ser así, la incertidumbre asociada a la calibración sería suficiente como para dar problemas a la hora de clasificar los espectros, ya que las bandas no coincidirán con los de la base de datos (o requerirán de una mayor tolerancia en el posicionamiento).

#### **2.3.1.2 Ajuste del rango espectral (recorte)**

Para mejorar el funcionamiento de los algoritmos de eliminación de la línea de base, es recomendable cortar del espectro aquellas zonas con caídas bruscas en la curva debidas al diseño óptico (filtros o fin de la huella de luz del espectro en el espectrómetro). Por tanto, se procede a recortar el espectro (figura 8) desde la zona donde termina la primera subida (filtro del cabezal que elimina la longitud de onda de excitación) hasta el punto donde empieza la bajada de éste (debido al final de rango espectral). Todo lo que



queda fuera de este rango no aporta información útil de ningún tipo y por tanto no es relevante para el análisis.

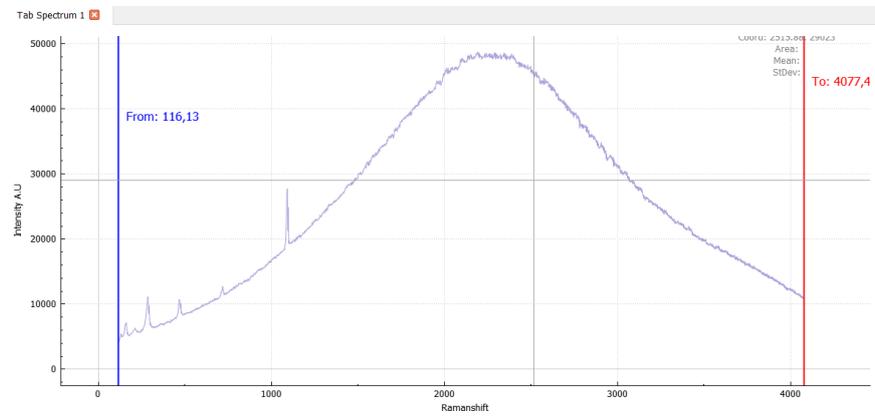


Figura 8. Espectro recortado

### 2.3.1.3 Eliminación de la línea de base

Uno de los pasos críticos en el tratamiento espectral es la eliminación de la línea de base del espectro para poder detectar mejor los picos presentes en el mismo. Para eliminar la línea de base se utiliza un algoritmo que depende de los parámetros recogidos en la Figura 9 y descritos a continuación [10]:

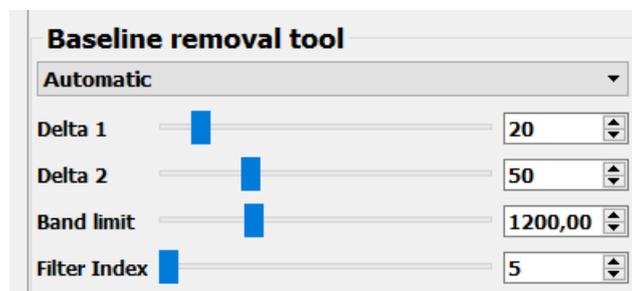


Figura 9. Parámetros para la eliminación de línea de base

- Delta 1: un valor cuya función es acercar la línea de base a la base del espectro problema.
- Band limit: a partir de este valor del eje de abscisas delta 1 deja de influir y empieza Delta 2.
- Delta 2: misma función que Delta 1.
- Filter index: el índice de un filtro de ventana deslizante que permite suavizar levemente el espectro para evitar cambios abruptos en la línea de base. Un valor 1 equivale a no hacer nada.

### 2.3.1.4 Normalización del espectro

Una vez eliminada la línea de base lo único que nos queda por hacer es la normalización del espectro ya que para que el algoritmo desarrollado no dependa de la magnitud de



los picos es conveniente utilizar espectros normalizados. De esta forma la intensidad del pico principal siempre es 1, facilitando la tarea de identificación del algoritmo. El método de normalización consiste en dividir todos los valores del espectro por la intensidad del pico más alto, quedando el espectro acotado a valores de intensidad entre 0 y 1.

### 2.3.1.5 Detección de picos y ajuste de bandas

Finalmente, el proceso de detección de picos y ajuste de bandas es el que proporcionará la información espectral necesaria para la ejecución del algoritmo. La herramienta de detección de picos de SpectPro permite identificar todos aquellos picos que se encuentran por encima de un determinado valor (que puede ser ajustado gráficamente), como se muestra en la figura 10. Una vez identificados los picos, la herramienta de ajuste de bandas genera automáticamente bandas de carácter gaussiano/lorentziano según la información obtenida por la herramienta de detección de picos (posición, intensidad y anchura) [10].

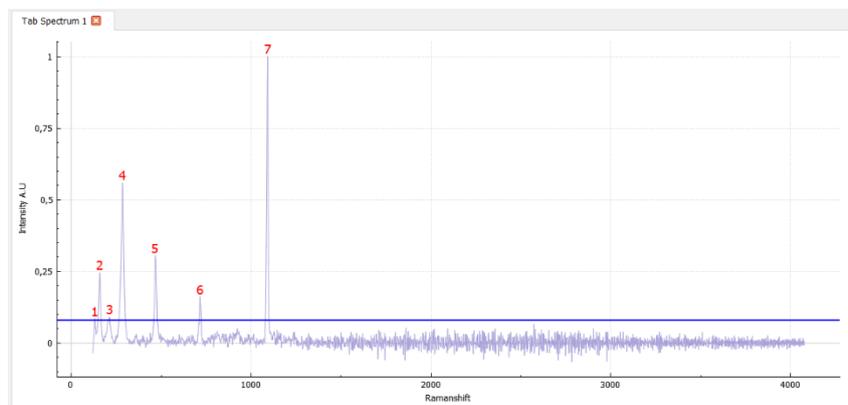


Figura 10. Bandas detectadas por el algoritmo en un espectro de calcita y cuarzo

Una vez hayamos identificado las bandas del espectro problema, éste se guarda generando un archivo xml (extensible markup language) de formato bds, donde se almacenan cada una de las bandas junto con su nombre para poder identificarla, intensidad normalizada, la posición, su anchura y el valor para el ajuste de banda según sea gaussiana/lorentziana, como se muestra en el ejemplo de la figura.

Un ejemplo de archivo xml se muestra en la figura 11. Se puede ver la estructura con la que se guardan los datos y todos los parámetros que se almacenan. Las distintas bandas se distinguen por el nombre de las etiquetas.



```
<?xml version="1.0" encoding="UTF-8"?>
<ListFittingBands>
  <Band>
    <bandName>Band 1</bandName>
    <bandPosition>155.50</bandPosition>
    <bandIntensity>0.22</bandIntensity>
    <bandFWHM>10.11</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 2</bandName>
    <bandPosition>282.50</bandPosition>
    <bandIntensity>0.49</bandIntensity>
    <bandFWHM>13.27</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 3</bandName>
    <bandPosition>712.50</bandPosition>
    <bandIntensity>0.15</bandIntensity>
    <bandFWHM>7.74</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 4</bandName>
    <bandPosition>1086.50</bandPosition>
    <bandIntensity>1.00</bandIntensity>
    <bandFWHM>8.34</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 5</bandName>
    <bandPosition>1436.00</bandPosition>
    <bandIntensity>0.03</bandIntensity>
    <bandFWHM>7.85</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 6</bandName>
    <bandPosition>1748.00</bandPosition>
    <bandIntensity>0.03</bandIntensity>
    <bandFWHM>9.19</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
</ListFittingBands>
```

Figura 11. Ejemplo de archivo xml

### 2.3.2 Creación de los ficheros de identificación

Aunque habitualmente un espectro que se utilice con los algoritmos de identificación será un espectro desconocido, o espectro problema, para realizar la correcta puesta a punto, testeo y parametrización de los algoritmos, se necesita trabajar con espectros conocidos. Por ello, para la identificación de los materiales de los que se compone el set de espectros problema utilizados en este trabajo (ver sección 2.1), se han creado una serie de ficheros txt para cada espectro, incluyendo el nombre de los materiales que los componen. Dichos ficheros se almacenan en la misma carpeta de los espectros problema correspondientes.

De esta forma, estos ficheros podrán ser utilizados por el algoritmo para conocer si la estimación realizada es correcta. Con esta metodología se facilita la automatización del proceso identificación de materiales del set de espectros problema, facilitando la ejecución por lotes del algoritmo, y por tanto una más fácil interpretación y parametrización del funcionamiento del algoritmo.

## 2.4 Propuesta de algoritmo de identificación automática



En este trabajo se ha propuesto un algoritmo para la identificación automática de espectros basado en la información de los picos (posición e intensidad) de los mismos. Si bien durante la ejecución en la fase de operación del instrumento RLS el procesamiento de los espectros para extraer los datos de los picos se realizará de forma automática, en este trabajo hemos centrado la actividad en el desarrollo de la rutina de identificación. Es decir, el desarrollo del algoritmo que, a partir de las posiciones e intensidades normalizadas de los picos (siempre el máximo valor de pico vale 1) proporcionará una mejor estimación del material o materiales que dan lugar a dicho espectro.

La propuesta del algoritmo se realiza en varias etapas y estableciendo un árbol de coincidencias de forma recursiva, para lo cual es necesario establecer dos parámetros de tolerancia en posición e intensidad que se utilizarán para establecer el criterio de coincidencia de picos (para poder considerar dos picos como coincidentes).

Por otro lado, de entre todas las posibles coincidencias será necesario realizar un proceso de clasificación. Dicha clasificación se ha propuesto en base a un criterio que considera tres aspectos: 1- la precisión en la posición de los picos, 2- la precisión en la intensidad de los picos, y 3- el porcentaje de acierto en número de picos del espectro problema frente a los espectros de la base de datos. Dichos criterios se utilizan para definir un único indicador gobernado por las siguientes ecuaciones:

$$\text{Coincidencia} = \left(1 - (P_{int} \cdot \text{Ind}_{int} \times + P_{pos} \cdot \text{Ind}_{pos} + P_{band} \cdot \text{Ind}_{band})\right) \times 100$$

$$P_{int} + P_{pos} + P_{band} = 1$$

Donde  $\text{Ind}_{int}$  es el indicador de coincidencia de intensidad, calculado como el promedio de las diferencias de todos los picos coincidentes del espectro;  $\text{Ind}_{pos}$  el indicador de coincidencia de las posiciones, calculado como el promedio de las diferencias entre las posiciones de los picos coincidentes;  $\text{Ind}_{band}$  el indicador de número de bandas coincidentes, en porcentaje con respecto al número de bandas coincidentes del espectro de referencia con más bandas coincidentes; y donde  $P_{int}, P_{pos}, P_{band}$  son los pesos aplicados a cada indicador.

Cuanto más parecido es el espectro de un material al espectro problema más pequeños son los indicadores de la diferencia de intensidad, la diferencia de posición y el valor de las bandas normalizadas. Para dar un valor de coincidencia en porcentaje estimado, se introduce el 1 en la ecuación, haciendo que la suma de indicadores se invierta y que el material más parecido tenga el valor más alto.

El funcionamiento del algoritmo es el siguiente:

### 1. Lectura de datos

Se leerán los parámetros (nombre, posición e intensidad) de los espectros de los minerales que se encuentran en la base de datos.



Se leerán esos mismos parámetros del espectro problema que queramos resolver.

## **2. Preselección de espectros**

Se buscarán todos aquellos minerales de la base de datos que tengan su pico principal (aquel pico cuya intensidad tiene valor 1) en el rango de posiciones donde esté el pico principal del espectro problema. Dicho rango viene determinado por el parámetro de la tolerancia de posición. Por ejemplo, si el pico principal del espectro problema se encuentra en la posición 450 y se ha dado una tolerancia de posición de 20, el rango será 430-470).

## **3. Cálculo de coincidencias**

De entre todos los espectros que se han preseleccionado como posibles candidatos (en base a las coincidencias del pico principal), se buscan las coincidencias del resto de picos.

Para ello se comparan las posiciones de las bandas del espectro problema con las bandas de cada material, uno a uno. Para cada banda, si ésta coincide, se comprueba si su intensidad también coincide (en el rango definido por la tolerancia de intensidad). Si la banda cumple los criterios de posición e intensidad se considera coincidente, y en ese caso se calculan las diferencias de posición y de intensidad. Este proceso se ejecuta para cada banda del espectro problema. El promedio de las diferencias obtenidas en posición e intensidad para todas las bandas, así como del número de bandas coincidentes se utiliza para calcular el valor de coincidencia según la fórmula descrita anteriormente. Para evitar sesgos debidos al número de bandas del espectro de referencia, se normaliza el número de bandas coincidentes con el espectro problema por el número de bandas coincidentes de aquel material que más haya obtenido.

## **4. Clasificación**

De entre todos los posibles materiales, se considera el más probable aquel que obtiene un mayor valor en base al indicador de coincidencia de la fórmula anteriormente descrita.

## **5. Actualización de la información de bandas del espectro problema**

Una vez hallado el material más probable, se resta a las bandas del espectro problema la intensidad de las bandas coincidentes de dicho material. Si la intensidad resultante de la resta es menor que la tolerancia de intensidad establecida, la banda se elimina. Si no, la intensidad se actualiza al valor de la diferencia. Es decir, se elimina la contribución estimada del espectro detectado en el espectro problema. Finalmente, el espectro se normaliza respecto al pico con más intensidad (éste se vuelve el pico principal).

## **6. Búsqueda de nuevos materiales**

Se comprueba si el espectro problema sigue teniendo bandas que no han sido utilizadas o eliminadas. Si es así, se vuelve al punto 3 hasta que se han procesado todas las bandas. Si por un casual, se vuelve al punto 3 y no se encuentra ningún material cuyas bandas coincidan con las bandas restantes, el bucle se terminaría



ya que la lista de posibles materiales estará vacía y se acabaría la búsqueda, así no tendremos problemas de bucles infinitos.

Los detalles de la implementación del algoritmo realizado, utilizando Matlab, se encuentran recogidos en el Anexo 1.

## 2.5 Filosofía de pruebas y parametrización del algoritmo

A la hora de la comparación del espectro problema con los espectros de la base de datos y clasificación de los materiales resultantes, se han definido un total de cinco parámetros tal y como se ha introducido en el apartado anterior (tolerancias en intensidad y posición para definir las coincidencias; y pesos de los criterios de clasificación en base a posición, intensidad y número de bandas).

Para encontrar la mejor combinación de todos estos parámetros, se ha adoptado una filosofía de análisis por fuerza bruta que consiste en analizar todos los espectros problema identificados en la sección 2.1, utilizando diferentes combinaciones de parámetros mediante bucles anidados. De esta forma se barrerán todas las posibles combinaciones de dichos parámetros en determinados rangos de valores. Posteriormente se analizarán los resultados buscando aquella o aquellas combinaciones de parámetros que resulten en un mejor funcionamiento del algoritmo.

El criterio para la definición del grado de bondad del algoritmo se basa asimismo en dos indicadores: 1- el porcentaje de aciertos con respecto al número total de materiales presentes en los espectros problema (cuanto mayor, mejor); y 2- el porcentaje de falsos positivos (es decir, materiales identificados erróneamente por el algoritmo) con respecto al número total de materiales identificados. Estos dos parámetros han de ser estudiados conjuntamente para encontrar la combinación de mayor bondad, si bien en general hemos aplicado el criterio de minimización del número de falsos positivos (identificar erróneamente algo puede ser más problemático que no identificar nada).

Para poder estudiar la gran cantidad de datos generados, se representan 2 tipos de gráficas: Una gráfica mostrará la tasa de aciertos, y la otra gráfica la tasa de falsos positivos. La representación consiste en una imagen bidimensional con los valores de tolerancias en los ejes x e y, para cada combinación de pesos de clasificación.

El barrido de parámetros de las tolerancias se realiza considerando las siguientes combinaciones, que se consideran más que razonables ya que abarcan un rango muy amplio de valores:

- A la tolerancia de intensidad se le dan valores entre 0,1 y 0,9 con incrementos de 0,2 (el rango total de valores posibles de intensidad varía entre 0 y 1 al estar los espectros normalizados)
- Con la tolerancia de posición se hace un barrido que se mueve entre 3 y 25 en pasos de 2 números de onda. En este caso se considera un rango más que suficiente teniendo en cuenta que diferentes fases minerales pueden tener



bandas separadas muy pocos números de onda, pero necesarios para proporcionar una cierta tolerancia o robustez para poder analizar espectros obtenidos con diferentes instrumentos.

En cambio, los pesos, para cada valor de los 2 parámetros anteriores, se combinan entre ellos de sin necesidad de hacer todos los posibles barridos, ya que la suma de pesos siempre tiene que ser igual a 1. Por tanto, se realiza de la siguiente manera:

**Peso de intensidad = 0:** El peso de posición varía de 0,1 a 0,9 en pasos de 0,2 mientras que el valor del peso de bandas hace el mismo barrido, pero en sentido contrario, de 0,9 a 0,1 en pasos de 0,2.

**Peso de intensidad = 0,2:** En este caso el valor del peso de posición se mueve de 0,1 a 0,7 en pasos de 0,2 y el valor del peso de bandas, como en el caso anterior, hace el mismo barrido en sentido contrario.

**Peso de intensidad = 0,4:** En este caso se prueban valores de 0,1 a 0,5 en pasos de 0,2 para el peso de posición y en el peso de intensidad es en sentido contrario.

**Peso de intensidad = 0,6:** En este caso el peso de posición se prueba con valores 0,1 y 0,3, y para el peso de intensidad se da 0,3 y 0,1.

**Peso de intensidad = 0,8:** Como última combinación de pesos, se le da un valor de 0,8 al peso de intensidad y a los otros 2 pesos se les da 0,1 a cada uno

Con este conjunto de parámetros se considera que es suficiente para hallar la mejor combinación que proporcione la mayor cantidad de aciertos y el menor número de falsos positivos, de esta forma pudiendo definir la parametrización que proporcionará el mejor comportamiento del algoritmo. El código implementado para la realización de la parametrización del algoritmo se realiza con Matlab, y está recogido en el Anexo 2.

### 3 Resultados y discusión

En este apartado analizaremos los resultados obtenidos al realizar el barrido por los diferentes valores de los parámetros, con el objetivo de encontrar la combinación con la que encontramos el mayor número de aciertos y el menor número de falsos positivos, con la filosofía recogida en apartados anteriores.

Para ello hemos dividido el total de espectros problema en 3 conjuntos diferentes. En primer lugar, se ejecuta el script de barrido de parámetros con, únicamente, los espectros puros. En segunda instancia se realiza para los espectros que contengan 2 materiales y, por último, la tercera ejecución se realiza sobre ambos tipos de espectros conjuntamente. El objetivo de esta división es estudiar si el algoritmo se comporta de forma óptima con los mismos parámetros para diferentes tipos de espectros, o por el contrario sería recomendable utilizar diferentes combinaciones de parámetros si existe algún tipo de información previa que permita conocer si hay uno o varios materiales en



dicho espectro. O también verificar si el algoritmo es capaz de identificar un material o varios en aquellos espectros con varios materiales, y de qué manera.

Para apoyar las explicaciones y estudiar en profundidad el algoritmo, se han incluido análisis detallados del funcionamiento para algunos espectros seleccionados, que se recogen en el Anexo 6.

### 3.1 Espectros puros

Los materiales de los espectros puros utilizados para la prueba están descritos en la tabla 2. Las gráficas resultantes del análisis se encuentran en el Anexo 3.

Como ya se ha explicado, las gráficas están compuestas por 2 figuras. La figura de la izquierda nos indica el porcentaje de aciertos que se obtienen para las diferentes combinaciones de tolerancias de intensidad y posición, para una determinada combinación de pesos del indicador (peso de posición, peso de bandas y peso de intensidad). Por otro lado, la figura de la derecha nos indica el porcentaje de falsos positivos que han surgido para esa combinación.

El objetivo principal de este proceso de parametrización es encontrar una combinación de parámetros que maximice el número de aciertos minimizando simultáneamente el número de falsos positivos sea pequeño.

Sin embargo, en seguida se observa que éste no es siempre el caso. Si analizamos la primera figura 31, para la combinación (Peso posición, Peso Bandas, Peso Intensidad) = (0,1, 0,9, 0), la mayor tasa de aciertos obtenida es del 100%. Ésta se obtiene en el rango 8-14 de las tolerancias de posición para el rango 0,2-1 de las tolerancias de intensidad; y en el rango 8-20 para las tolerancias de intensidad entre 0 y 0,2. Por otro lado, el menor porcentaje de falsos positivos (25%) se encuentra en el rango 4-6 de tolerancias de posición y 0-0,2 de tolerancias de intensidad. Por tanto, el mayor porcentaje de aciertos y el menor porcentaje de falsos positivos no coinciden para los mismos rangos en este caso.

El estudio detallado de todos los casos muestra que existen combinaciones en las que sí se consigue que coincida el máximo número de aciertos con el mínimo número de falsos positivos. Por ejemplo, la figura 43 muestra los resultados obtenidos con la combinación de pesos = [0,1-0,3-0,6]. El máximo de aciertos y el mínimo de falsos positivos se encuentra en la misma franja: rango 8-16 de tolerancias de posición y 0,2-1 de tolerancias de intensidad, pero el porcentaje mínimo de falsos positivos es de 30% al igual que en el caso en la figura 31 por lo que, al fin y al cabo, estamos obteniendo un resultado equivalente.

El mínimo absoluto de falsos positivos que podemos encontrar entre todas las combinaciones es del 25%. Sin embargo, con esa tasa obtenemos un 85% de aciertos. No obstante, existen combinaciones que sí que proporcionan una tasa de aciertos en los



minerales puros del 100%, con una tasa de falsos positivos del 30%. En estos casos, al detectar el 100% de materiales (que, recordemos, son puros), la conclusión es que el 30% de falsos positivos son debidos a que el algoritmo considera que en un espectro hay más de un material.

Esto se produce cuando el espectro problema tiene más picos que el espectro de referencia, o bien cuando la tolerancia en intensidad sea pequeña y la diferencia entre las intensidades de uno y otro espectro sean muy grandes. En este sentido, se ha observado que la diferencia en intensidades relativas entre picos de los espectros problema y los de la base de datos es, en muchos casos, muy acusada. Esto es debido a la respuesta instrumental de los diferentes instrumentos utilizados para la adquisición de los espectros. En este sentido, una correcta calibración en intensidad (pero también en posición) de los espectros problema, facilitaría la detección por parte del algoritmo. No obstante, para solucionar esta situación en el algoritmo, se podría buscar alguna modificación del mismo para eliminar de forma definitiva una banda cuando haya habido una coincidencia (independientemente de su intensidad). Esto simplificaría, quizá eficazmente, el algoritmo (elimina la dependencia con la intensidad), si bien elimina también una de sus grandes potencialidades (la detección múltiple de espectros con bandas coincidentes).

Otro problema del diseño del algoritmo estaría relacionado con la definición de los indicadores, como se puede ver en el análisis detallado del anexo 6, sección 5.6.1, donde se muestra cómo espectros con una única banda coincidente obtienen un mejor indicador de posición que otras que tienen varias bandas coincidentes. Una posible forma de mejorar el algoritmo podría ser **reducir el número de parámetros, incluyendo el número de bandas como multiplicador implícito de los indicadores de posición e intensidad**. Esto reduciría el número de parámetros por un lado simplificando el algoritmo, mientras que podría mejorar los indicadores de posición e intensidad.

Por otro lado, si existe una diferencia en el número de bandas de los espectros, el problema es metodológico: o bien la base de datos tiene un error y le faltan bandas, o bien el método de extracción de la información espectral del espectro problema es errónea e incluye más bandas de las realmente existentes. Un ejemplo de esta casuística se muestra en el anexo 6, sección 5.6.2. En todo caso, este análisis muestra la importancia de realizar un correcto análisis espectral, tanto de los espectros de la base de datos, como de los espectros problema. Teniendo en cuenta que, durante la operación de RLS, la extracción de los parámetros espectrales se realizará de forma autónoma, no es descartable pensar en **desarrollar algoritmos que sean robustos frente a posibles errores en los picos de los espectros problema**.

La mejor combinación para estos espectros la encontramos cuando los pesos tienen un valor de 0,3 el de posición, 0,3 el de las bandas y 0,4 el de las intensidades, tal y como se muestra en la figura 12.

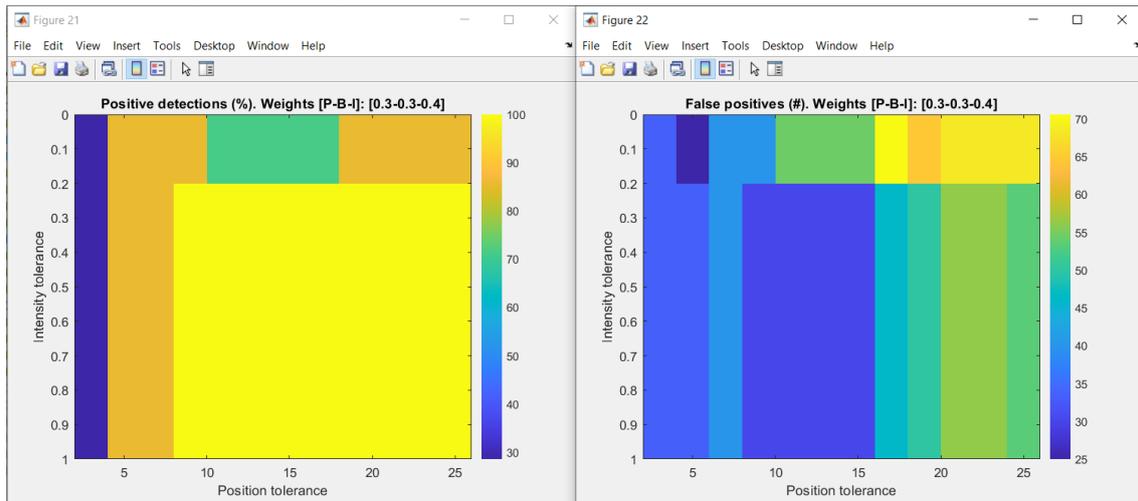


Figura 12. [P-B-I] = [0.3-0.3-0.4]

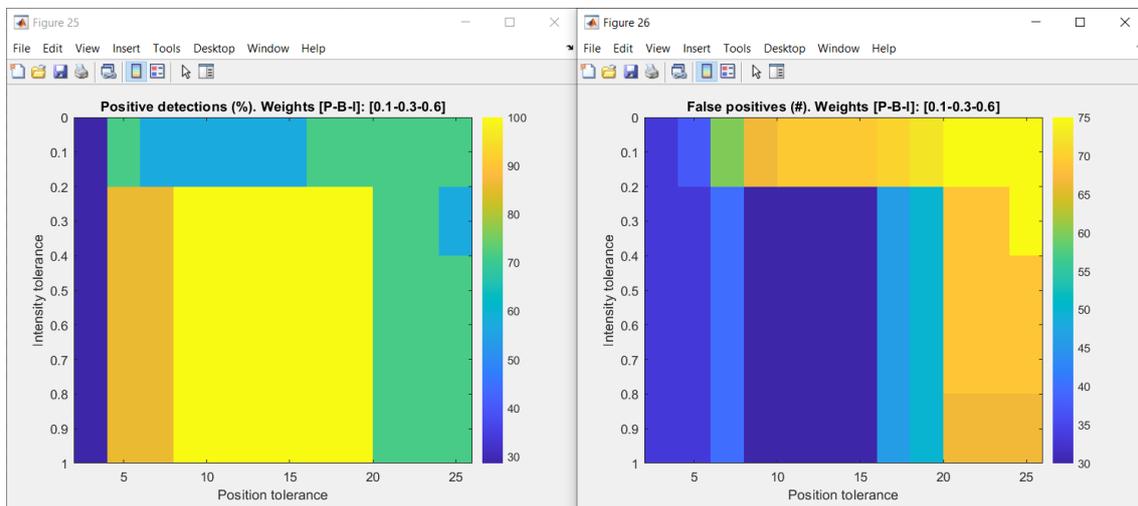


Figura 13. [P-B-I] = [0.1-0.3-0.6]

Es una combinación que nos permite una tasa de aciertos del 100% y una tasa de falsos positivos del 30% en el rango de tolerancias de posición 8-16 y en las tolerancias de intensidad mayores que 0,2. El valor de las variables no llega a tener un valor extremo. Con una tolerancia de posición de 10 y una tolerancia de intensidad de 0,4 obtendríamos un 100% de aciertos. Sin embargo, hay otras combinaciones (por ejemplo [0,1-0,3-0,6], ver figura 13) que, comportándose de manera similar para espectros puros, podrían obtener mejores resultados considerando también los espectros de mezclas, como se muestra en las siguientes subsecciones.

### 3.2 Espectros mezcla

Al haber observado una incidencia elevada de falsos positivos con los espectros puros, se muestra cómo el algoritmo favorece la búsqueda de materiales secundarios en el espectro (lo que por otro lado es habitual, ya que no suelen encontrarse espectros de



materiales puros en la naturaleza, especialmente en muestras en polvo y homogeneizadas como en el caso de ExoMars).

Con el análisis sobre los espectros mezcla de varios materiales (tabla 3), se comprueba que hay una tasa menor de aciertos, pero también de falsos positivos. Por ejemplo, cuando se obtiene el máximo de aciertos (82%), la tasa de falsos positivos se encuentra entorno al 20%.

Sin embargo, estudiando en detalle todos los casos, se observa cómo en las figuras 14 (0,1-0,5-0,4) y 15 (0,1-0,3-0,6) se obtiene una tasa de falsos positivos muy baja, rondando el 5%, para una tasa de aciertos del 75%. Por ello, en general, parece óptimo seleccionar una de estas combinaciones.

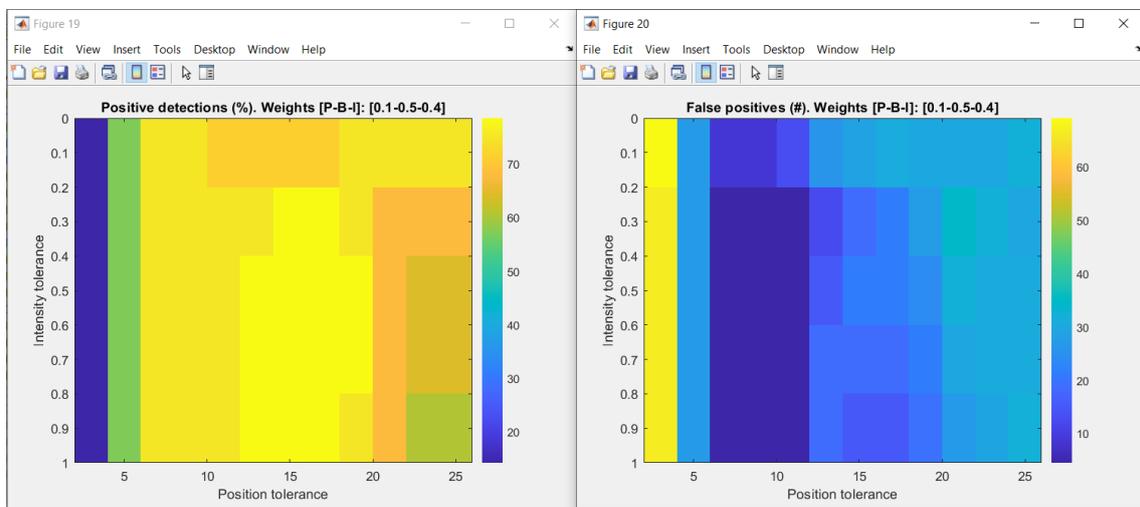


Figura 14. [P-B-I] = [0.1-0.5-0.4]

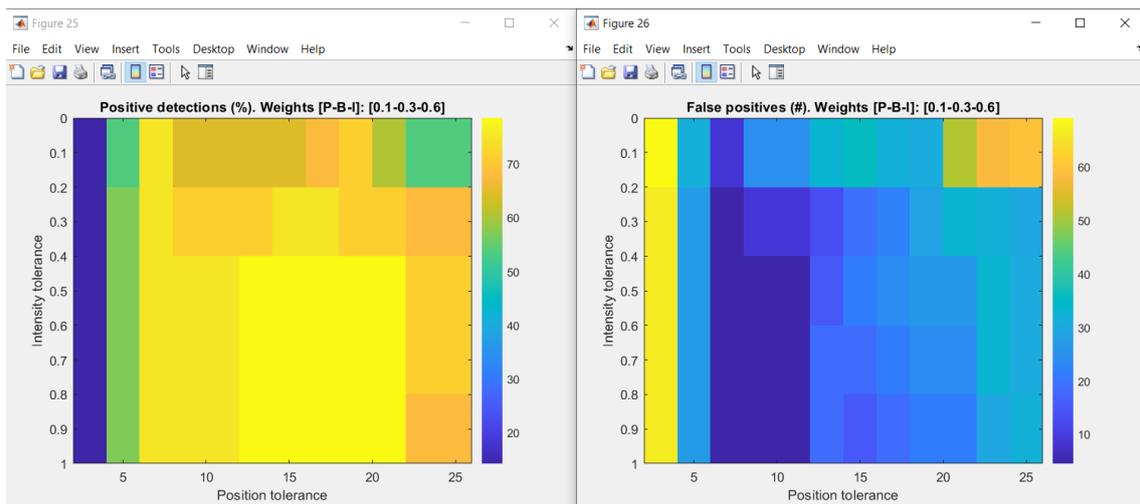


Figura 15. [P-B-I] = [0.1-0.3-0.6]

De hecho, comparando con los datos obtenidos con los espectros puros, se puede obtener un 100% de acierto en espectros puros, y un resultado muy bueno en mezclas binarias. Es más, un análisis en profundidad de los casos [0,3-0,3-0,4] y [0,1-0,3-0,6],



muestra cómo en el 100% de los casos se detecta el elemento mayoritario de las mezclas, si bien el segundo caso obtiene un resultado muy superior en términos de falsos positivos (5% vs 30%). En todo caso, todos los falsos positivos se obtienen en la detección de los materiales secundarios. Un caso concreto que demuestra cómo se producen los falsos positivos se ha detallado en el Anexo 6, sección 5.6.3.

También se ha podido volver a observar cómo la ausencia de materiales en la base de datos es muy crítica a la hora de obtener una mayor incidencia de falsos positivos. La ejecución del algoritmo eliminando los espectros de olivino+serpentina (ya que la serpentina no se encuentra en la base de datos, y el espectro problema tenía como elemento mayoritario la serpentina) proporcionó mejoras sustanciales en el porcentaje de falsos positivos y detecciones correctas (incrementos de hasta un 10% en términos absolutos). **De cara a futuro, sería necesario eliminar aquellos picos principales que no se detecten en la base de datos**, favoreciendo la detección de materiales minoritarios y mejorando así la robustez del algoritmo. Por supuesto, utilizar una base de datos completa, así como de espectros problema correctamente calibrados (en posición e intensidad) mejoraría notablemente el funcionamiento del algoritmo.

Considerando conjuntamente los espectros puros y de mezclas, se concluye que el algoritmo es capaz de detectar el elemento mayoritario (siempre que se encuentre en la base de datos y que la extracción de parámetros sea correcta) en un 100% de las ocasiones, mientras que detecta el elemento secundario con una fiabilidad del 70% (ya que el 30% es la incertidumbre obtenida para los materiales puros). Detecciones subsiguientes de minerales en mezclas con más materiales no han sido analizadas en detalle, pero es presumible que la fiabilidad vaya decreciendo en cada iteración.

### 3.3 Combinación de todos los espectros

Al realizar el barrido de los parámetros de forma conjunta todos los espectros, se obtienen los resultados que esperábamos al analizar los 2 apartados anteriores, obteniendo valores intermedios a los obtenidos en los espectros puros y mezclas por separado. El porcentaje de aciertos disminuye respecto a la prueba con los espectros puros, llega a un 86% como máximo. El porcentaje de falsos positivos mínimo llega a 15%.

Analizando los resultados de este barrido, parece que una buena combinación de parámetros es la que se muestra en la figura 16, donde la tolerancia de posición tiene un valor centrado en 10, la tolerancia de intensidad tiene que ser mayor de 0,4, y la combinación de pesos es [0,1-0,3-0,6]. Con esta combinación tenemos una tasa de aciertos mayor que 80% y una tasa de falsos positivos menor que el 15%. Además, esta combinación es coherente con los resultados obtenidos para los diferentes tipos de espectros problema por separado, según se ha visto anteriormente.

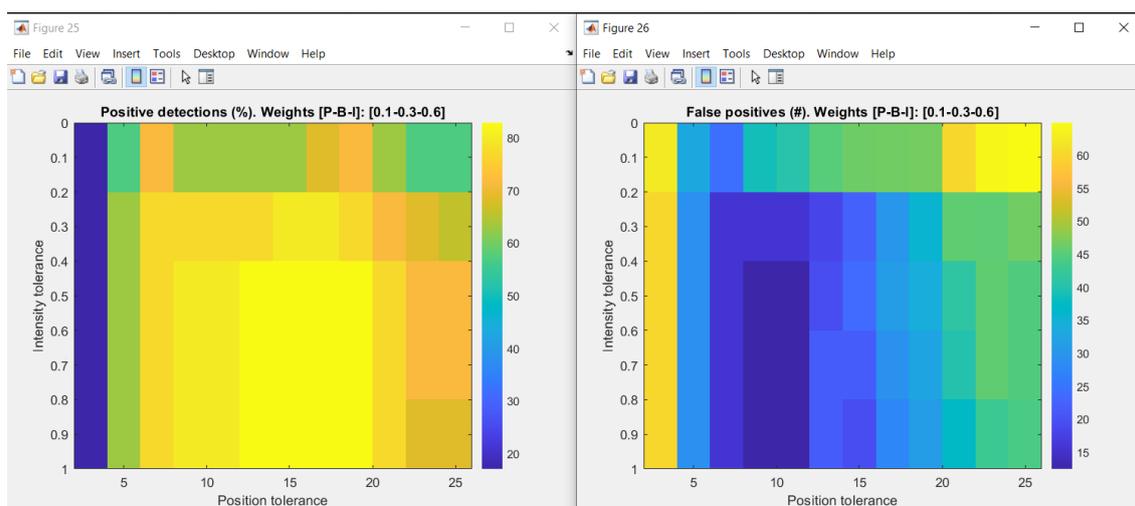


Figura 16. [P-B-I] = [0.1-0.3-0.6]

## 4 Conclusiones y líneas futuras

El objetivo del trabajo ha sido desarrollar un algoritmo para la identificación de las fases minerales presentes en espectros Raman.

El trabajo se ha dividido en 3 partes:

1. La primera parte del trabajo ha consistido en generar los archivos xml donde se incluyen los valores de los parámetros de nombre, posición e intensidad de todos los espectros problema de los que se ha tenido acceso usando el software SpectPro.
2. La segunda parte se basó en la creación del algoritmo para la identificación de las fases minerales.
3. Por último, se buscó la mejor combinación de los parámetros para utilizar el algoritmo principal.

Después de probar el código en diferentes situaciones y realizar un barrido de fuerza bruta probando múltiples combinaciones de parámetros en diferentes conjuntos de espectros (puros, mezclas y todos en conjunto), se ha llegado a la conclusión de que es un algoritmo potencialmente útil para cumplir su finalidad prevista, ya que existe al menos una combinación de parámetros  $[P-B-I] = [0,1-0,3-0,6]$  y tolerancias  $[pos-int]=[10-0,4]$  tal que, siempre que un material se encuentre en la base de datos, el algoritmo es capaz de detectar el elemento mayoritario (en espectros puros y de mezclas). Además, el algoritmo ha demostrado ser capaz de detectar el elemento secundario presente en mezclas binarias con un 70% de probabilidades (hay un 30% de posibilidades de obtener un falso positivo).

Como líneas futuras en este sentido, se cree haber identificado el posible problema relacionado con los falsos positivos en los espectros puros, por lo que se han propuesto



pequeñas modificaciones y/o simplificaciones al algoritmo para intentar mejorar dichas prestaciones y reducir la incertidumbre en la detección de los materiales minoritarios.

Por otro lado, se ha podido comprobar que algunos de los fallos que se han producido eran debidos a la extracción de datos espectrales de los espectros de la base de datos o los espectros problema. Éstos han dado muchos problemas debido a que algunos de ellos no estaban del todo calibrados (lo cual puede tener un efecto en las tolerancias óptimas), o a la hora de guardar sus bandas en el archivo xml, éstas se guardaban incorrectamente favoreciendo la generación de falsos positivos. Por tanto, es muy importante trabajar con espectros correctamente procesados. De todos modos, durante la operación del instrumento RLS el procesamiento de los espectros para extraer los datos se realizará, al menos durante las primeras horas después de la recepción de los datos, de forma automática (y por tanto potencialmente más inexacta), por lo que se ha considerado más interesante plantear modificaciones que mejoren la robustez del algoritmo.

Nuevas líneas de investigación podrían incluir tecnologías como son las redes neuronales. Para ello, primero habría que estudiar los espectros de la base de datos para ver cómo estructurar la red. Estudiar dónde se encuentran los picos de las bandas de los espectros y si fuera mejor separar éstos por tipos (carbonatos, sulfatos, óxidos, materiales hidratados) y ejecutar en varias etapas (emulando lo que de alguna manera hace el algoritmo en este trabajo, al clasificar en primer lugar el pico principal), o directamente una única red capaz de clasificar todos los minerales.



## 5 Anexos

### 5.1 Anexo 1: Descripción del algoritmo

#### 5.1.1 Desarrollo de rutina de identificación

El algoritmo desarrollado en este trabajo está hecho en su totalidad con el entorno Matlab.

La rutina no solo tiene que identificar el compuesto principal del espectro, sino que, si hay algún otro compuesto en el espectro también tiene que lograr adivinar cual es.

#### 5.1.2 Matlab

La elección del entorno de Matlab para el desarrollo de la rutina se fundamenta en que este lenguaje se utiliza para: aprendizaje automático, procesamiento de señales, procesamiento de imágenes, visión artificial, comunicaciones, finanzas computacionales, diseño de control, robótica etc.

El entorno Matlab está optimizado para resolver problemas científicos y de ingeniería. Las gráficas integradas en el lenguaje facilitan la visualización de los datos y la obtención de información de ellos. También, tiene integrado una biblioteca de herramientas (Toolbox) que permiten empezar a trabajar inmediatamente con algoritmos esenciales.

Las características principales son:

- Lenguaje de alto nivel para cálculos científicos y de ingeniería.
- Entorno de escritorio optimizado para la exploración iterativa, el diseño y la solución de problemas.
- Gráficas para visualizar datos.
- Aplicaciones para ajustar curvas, clasificar datos, analizar señales, ajustar sistemas de control y muchas otras tareas.
- Toolbox complementaria para una amplia variedad de aplicaciones científicas.

Por estas razones, se ha optado por el lenguaje de programación Matlab para el desarrollo de los scripts.

Del archivo xml que extraemos de los espectros, guardamos, de cada banda, su nombre para identificar la banda, su intensidad normalizada, su posición, su anchura y el valor para el ajuste de banda según sea gaussiana/lorentziana. Para crear la rutina de identificación de las bandas se ha decidido utilizar solo 2 parámetros, estos son, la posición y la intensidad.

Se han creado en total 6 scripts para la identificación de los materiales. Un primer script para la lectura y extracción de la información de los materiales de la base de datos, para esto se apoyará en una función que se desarrolla en otro archivo llamada "obtener\_datos\_fn". Un segundo script para la identificación de los materiales. 2 scripts que ayudarán al segundo script en su función de identificación. Y, por último, se usará



otro script para la visualización de las gráficas que nos darán información sobre que parámetros son mejores para la identificación de materiales.

### 5.1.3 Lectura de la base de datos

Para el script de la extracción de datos hemos usado dos bucles donde en el primer bucle se recorre todas las carpetas de la base de datos y en el segundo bucle se buscan los archivos cuya extensión es “bds”. Estos archivos son aquellos archivos xml donde se guarda la información de las bandas.

En la celda que se crea se guarda el nombre del material, el número de bandas que tiene y los datos de cada banda.

En total, se guardan 53 elementos.

El código es el siguiente:

```
path = '../Spectra/Standard_DB/';
Carpetas_materiales = dir(path);

Num_carpetas = size(Carpetas_materiales)-2;
Muestras = cell(Num_carpetas);

Num_Muestras = 0;

for i = 1:Num_carpetas
    pathname = fullfile(path, Carpetas_materiales(i+2).name);
    Espectros = dir(pathname);

    for z = 3:size(Espectros)
        k = strfind(Espectros(z).name, '.bds');
        if(k >= 0)
            Num_Muestras = Num_Muestras + 1;
            Muestras{Num_Muestras}.name = Carpetas_materiales(i+2).name;
            fileid = fopen([pathname '/' Espectros(z).name]);
            metadata=textscan(fileid,'%[\n]');
            fclose(fileid);
            Muestras{Num_Muestras}.NofBand = Obtener_NofBand_fn(metadata{1}, 'bandName');

            for j = Muestras{Num_Muestras}.NofBand:-1:1
                Muestras{Num_Muestras}.bands(j).name = obtener_datos_fn(metadata{1}, 'bandName', false, j);
                Muestras{Num_Muestras}.bands(j).posicion = obtener_datos_fn(metadata{1}, 'bandPosition', true, j);
                Muestras{Num_Muestras}.bands(j).intensidad = obtener_datos_fn(metadata{1}, 'bandIntensity', true, j);
                Muestras{Num_Muestras}.bands(j).FWHM = obtener_datos_fn(metadata{1}, 'bandFWHM', true, j);
                Muestras{Num_Muestras}.bands(j).Lorentzian = obtener_datos_fn(metadata{1}, 'lorentzianFlavour', true, j);
                Muestras{Num_Muestras}.bands(j).Gaussian = obtener_datos_fn(metadata{1}, 'gaussianFlavour', true, j);
            end
        end
    end
end
```

Figura 17. Script A

Para la lectura de los datos, se hace uso de la función “obtener\_datos\_fn”. Definida en la figura 18.





La primera tarea que hacemos para detectar el material del espectro es ir comprobando que los materiales de la base de datos que hemos recogido anteriormente tengan el pico principal en una posición que esté cerca de la posición del pico principal del espectro problema. Como puede haber pequeñas variaciones en la posición al recoger el espectro, al comparar las posiciones del espectro problema con las posiciones de los espectros de la base de datos, se añade una pequeña tolerancia de posición para crear un rango de posiciones.

Todos los materiales que cumplan que su pico principal se encuentre cerca de la posición del pico principal del espectro problema se guardan en una lista de posibles materiales.

```
lista = [];  
%%Comparación del espectro problema con la base de datos  
for i = Num_Muestras:-1:1 % Recorro materiales.  
    if(prueba(Muestras{i}, Espectro, tolerancia_posicion))  
        lista = [lista, Muestras{i}];  
    end  
end
```

Figura 20. Creación de la lista con posibles materiales

```
function matches = prueba(materialA, materialB, tolerancia_posicion)  
    matches = false;  
    materialB.bands = sortrows(struct2table(materialB.bands), 'intensidad', 'descend');  
    materialB.bands = table2struct(materialB.bands);  
  
    index = find([materialA.bands.intensidad] == 1);  
  
    if(index > 0)  
        if(abs(materialA.bands(index).posicion - materialB.bands(1).posicion) <= tolerancia_posicion)  
            matches = true;  
        end  
    end  
end
```

Figura 21. Código de la función para la creación de la lista de posibles materiales

### 5.1.5.2 Orden de la lista

El siguiente paso, después de haber creado la lista de posibles materiales, es ordenar esta lista según se parezcan más o menos sus las bandas a las bandas del espectro problema.



```
if isempty(lista)
    break;
end

for i = length(lista):-1:1 % Recorro materiales.
    coincidence = coincidence_calculator_fn(lista(i), Espectro, tolerancia_intensidad, tolerancia_posicion);
    lista(i).nBandas = coincidence(3);
    lista(i).intensidad = coincidence(1)*peso_intensidad;
    lista(i).posicion = coincidence(2)*peso_posicion;
end

A = max([lista.nBandas]);

for i = length(lista):-1:1
    lista(i).nBandas = 1-(lista(i).nBandas/A);
    lista(i).bandas = lista(i).nBandas*peso_bandas;
    lista(i).coincidence = (1-(lista(i).intensidad+lista(i).posicion + lista(i).bandas))*100;
end

lista = sortrows(struct2table(lista), 'coincidence', 'descend');
display(lista)
lista = table2struct(lista);
```

Figura 22. Orden de la lista

Para ordenar la lista lo que hago es pasar estos posibles materiales a la función “coincidence\_calculator\_fn” donde comprobamos si las demás bandas de esos materiales coinciden con las del espectro problema. Para ello, volvemos a hacer uso de la tolerancia de posición para detectarles.

Para usar la función creamos un bucle para ir enviando solo un material en vez de la lista entera.

Si alguna banda de un posible material coincide que se encuentra en el rango de posiciones dado, se procede a comprobar que su intensidad se parezca a la del espectro problema, ya que una situación probable a ocurrir es que la posición de una banda de un posible material coincida con la posición de la banda del espectro problema, pero entre sus intensidades haya mucha diferencia.

Si también cumple con la intensidad quiere decir que esta banda es correcta por lo que este material tiene una banda que coincide con la del espectro problema.

De esta banda, guardamos la diferencia de intensidad entre la banda del material y la banda del espectro problema, la diferencia entre sus posiciones y a la variable “nBandas” (número de bandas) le sumamos uno.

Si esta banda no cumple con la intensidad que se pide, a este material se le añade a su vector de diferencia de intensidades y a su vector de posiciones un 1 para que, al ordenar la lista, éste tenga una posición mala.

Acto seguido, volvemos a comprobar las demás bandas.



```
function coincidence = coincidence_calculator_fn(materialA, materialB, tolerancia_intensidad, tolerancia_posicion)
    intensidad = [];
    posicion = [];
    nBandas = 0;
    for j = 1:length(materialB.bandas) % Recorro bandas
        index = find((materialA.bandas.posicion) <= materialB.bandas(j).posicion + tolerancia_posicion & (materialA.bandas.posicion) >= materialB.bandas(j).posicion - tolerancia_posicion);
        if(index > 0)
            if(length(materialA.bandas(index).posicion) > 1)
                A = abs(materialA.bandas(index).posicion - materialB.bandas(j).posicion);
                [E, indice] = min(A);
                index = index(indice);
            end
            if(abs(materialA.bandas(index).intensidad - materialB.bandas(j).intensidad) <= tolerancia_intensidad)
                intensidad = [intensidad, abs(materialA.bandas(index).intensidad - materialB.bandas(j).intensidad)];
                posicion = [posicion, abs(materialA.bandas(index).posicion - materialB.bandas(j).posicion)/tolerancia_posicion];
                nBandas = nBandas + 1;
            else
                intensidad = [intensidad, 1];
                posicion = [posicion, 1];
            end
        end
    end
    coincidence = [mean(intensidad), mean(posicion), nBandas];
end
```

Figura 23. Código de la función de coincidencias

Un problema que ha surgido al comprobar el código ha sido que algunos materiales cumplen que 2 de sus bandas están cerca de una banda del espectro problema por lo que para solucionar esto seleccionamos la banda más cercana.

Esta función devuelve la media de la diferencia de intensidades y posiciones entre las bandas que cumplen con los requisitos y, además, el número de bandas que coinciden del material con las del espectro.

Una vez salimos de la función, los parámetros que se devuelven se guardan junto a su material.

Una vez pasados todos los materiales, se procede a buscar cual es el material que haya tenido más bandas parecidas. El valor de las bandas coincidentes de todos los materiales posibles es dividido por éste. Después, restamos ese valor de un 1, así, el material con el mayor número de bandas parecidas pasa a tener un 0 en este campo y los demás sus respectivos valores. Esto lo hacemos debido a que para los 3 parámetros que utilizamos para ordenar la lista (media de diferencia de intensidades, media de diferencia de posiciones y el número de bandas), cuan más pequeños sean, más se parece este material al espectro problema.

Al calcular estos 3 parámetros, el siguiente paso es multiplicar cada uno de ellos por un valor (peso). Este valor se divide en 3: peso\_intensidad, peso\_bandas y peso\_posicion. Los 3 pesos tienen que sumar 1. Al parámetro que queramos darle más valor le damos a su peso más valor.

Finalmente, para ordenar la lista, restamos a un 1 la suma de los valores de los parámetros después de la multiplicación y multiplicamos por 100 para tener un porcentaje.

Después de todo estos pasos, cada material tiene un campo donde se guarda un porcentaje que indica cuanto de parecido es al espectro problema. Solo nos queda ordenar la lista de mayor a menor. El primer material de la lista es el material principal del espectro problema.

### 5.1.5.3 Detección de un posible segundo material en el espectro



Una vez hayamos detectado el primer material, tenemos que asegurarnos de si el espectro problema sigue teniendo bandas que no hayan sido utilizadas para detectar el primer material.

Para ello, tenemos que eliminar aquellas bandas que han sido utilizadas para detectar el primer material. El código consiste en buscar aquellas bandas del primer material que coincidan con el del espectro problema.

```
for i = length(Espectro.bands):-1:1
    index = find((([lista(1).bands_posicion] <= Espectro.bands(i).posicion + tolerancia_posicion) & ([lista(1).bands_posicion] >= Espectro.bands(i).posicion - tolerancia_posicion)));
    if(index > 0)
        if(length([lista(1).bands(index).posicion]) > 1)
            A = abs([lista(1).bands(index).posicion] - Espectro.bands(i).posicion);
            [F, indice] = min(A);
            index = index(indice);
        end
        Espectro.bands(i).intensidad = abs(Espectro.bands(i).intensidad - lista(1).bands(index).intensidad);
        if(Espectro.bands(i).intensidad < tolerancia_intensidad)
            Espectro.bands(i) = [];
        end
    end
end
end

if isempty(Espectro.bands)
    break;
end
```

Figura 24. Eliminación de bandas

Si el espectro se queda vacío, significa que ya hemos utilizado todas las bandas del espectro para detectar los materiales por lo tanto nos salimos del bucle, si no, repetir todo el proceso anterior.

Después de finalizar el bucle, el script es capaz de decirnos que materiales son los del espectro distinguiendo todos los que tenga. Si es un espectro problema puro, nos sacaría solo uno mientras que, si hay una mezcla, nos saldría como resultado el material primario y todos los demás.

## 5.2 Anexo 2 - Verificación del script

Según se explica en la sección 2.5, al haber 5 parámetros de los cuales depende el código del anterior script para encontrar los materiales del espectro (tolerancia de intensidades, tolerancia de posiciones, peso de intensidad, peso de posición y peso de bandas) y según como los variemos, podemos acertar más o menos y tener la posibilidad de sacar falsos positivos, se ha hecho un script para poder encontrar los valores óptimos para estos parámetros. El objetivo es encontrar la combinación de parámetros que proporciona un número de aciertos grande con el mínimo número de falsos positivos.

Para comprobar esto, se ha creado un script donde se le dan diferentes valores a estos 5 parámetros para buscar su valor óptimo. Como resultado se implementan una serie de bucles anidados para resolver todos los espectros problema, con todas las combinaciones de parámetros. Por tratarse de ejecuciones de espectros individuales que llaman a funciones, no es posible utilizar indización y se requieren bucles anidados.

Se ha creado una carpeta donde se han guardado todos los espectros problemas, cada uno de ellos con un archivo txt donde se ha escrito los materiales que forman el espectro.



En la primera parte del código se llama al primer script para leer todos los espectros de la base de datos y se guardan los archivos xml de los espectros problemas de la carpeta donde se encuentran todos.

```
a_extract_bds;  
  
%Lectura de los datos del espectro problema  
path = './Espectros';  
nombres_espectros = dir(path);
```

Figura 25. Lectura de espectros de la base de datos

En la segunda parte del código se crea un bucle “while” para poder trabajar con todos los espectros problema. En este bucle, lo primero que se hace es leer el archivo xml del primer espectro, después se lee su archivo txt para saber que materiales forman ese espectro.

```
pathname_bds = fullfile([path '\' nombres_espectros(t).name]);  
fileid = fopen(pathname_bds);  
metadata=textscan(fileid, '%[^\n]');  
fclose(fileid);  
  
t = t+1;  
  
%Añadimos un número para leer el archivo txt  
pathname_txt = fullfile([path '\' nombres_espectros(t).name]);  
fileid = fopen(pathname_txt);  
data = textscan(fileid, '%s %[^\n]');  
fclose(fileid);
```

Figura 26. Lectura de archivo xml y archivo txt

Se dan valores a los 5 parámetros. En este caso, la tolerancia de intensidad se ha dado un rango de valores entre 0,1 y 0,9. La tolerancia de posición se le ha dado un rango de 3 a 25.

Referente a los pesos, se han creado 3 rangos diferentes y otros 3 valores para ellos. En el primer rango el peso de intensidad es 0, el peso de posición se mueve entre los valores 0,1 y 0,9 mientras que el peso de bandas va de 0,9 a 0,1. La suma de los 3 pesos tiene que valer siempre 1.

En el segundo rango, el peso de intensidad pasa a ser de 0,2, el peso de posición se mueve de 0,1 a 0,7 y el peso de bandas se mueve entre 0,7 y 0,1.



En el tercer rango, el peso de intensidad pasa a valer 0,4, el peso de posición está en un rango de 0,1 a 0,5 y el peso de bandas se mueve entre 0,5 y 0,1.

El cuarto valor que pasan a tener es de: el peso de intensidad vale 0,6, el peso de posición vale 0,1 y el peso de bandas 0,3. Luego se intercambian los valores de los pesos de posición y bandas.

Por último, el peso de intensidad es de 0,8 y el peso de posición vale 0,1 y el peso de bandas también vale 0,1.

```
tolerancia_intensidad = [0.1:0.2:0.9];  
tolerancia_posicion = [3:2:25];  
peso_posicion = [0.1:0.2:0.9 0.1:0.2:0.7 0.1:0.2:0.5 0.1 0.3 0.1];  
peso_bandas = [0.9:-0.2:0.1 0.7:-0.2:0.1 0.5:-0.2:0.1 0.3 0.1 0.1];
```

Figura 27. Valor de parámetros

Para poder probar todos los valores de los parámetros se ha creado una función (“c\_normalizacion\_fn”) que realiza la misma función que el segundo script cuyos argumentos de entrada son el espectro problema, los materiales de la base de datos y el valor de los 5 parámetros que toque en su momento.

Esta función devuelve un vector en el que se guardan los materiales que se han encontrado en el espectro y los valores de los 5 parámetros utilizados.

Este vector devuelto se guarda en una matriz de 4 dimensiones.

Después de establecer todos los valores necesarios para el funcionamiento del código, se procede a comprobar cuál es la combinación óptima de los 5 parámetros. Para ello se han creado 3 bucles “for” consecutivos donde se recorren los rangos que se han mencionado anteriormente.

Para comprobar que combinación de parámetros es la óptima se han creado 2 matrices “Primer\_indicador” y “Segundo\_indicador” de 4 dimensiones.

En la primera matriz se guarda el número de materiales acertados devueltos por la función con los valores de los 5 parámetros dados. Para ello, se comparan los materiales devueltos por la función con los materiales que se habían escrito en el archivo txt.

En la segunda matriz se guarda el número de falsos positivos que se han podido generar al elegir unos determinados valores de los parámetros como la resta entre todos los materiales detectados y los materiales acertados



```
for i = numel(tolerancia_intensidad):-1:1
    for j = numel(tolerancia_posicion):-1:1
        for z = numel(peso_posicion):-1:1
            disp(['Iter [i, j, z]: [' num2str(i) ', ' num2str(j) ', ' num2str(z) ']])
            materiales = 0;
            materiales_acertados = 0;
            peso_intensidad(z) = 1-(peso_posicion(z) + peso_bandas(z));
            lista(i, j, z, p) = c_normalizacion_fn(Espectro, Muestras, tolerancia_intensidad(i), tolerancia_posicion(j), peso_intensidad(z), peso_posicion(z), peso_bandas(z));
            dat_comp = data;
            if(iscell(lista(i, j, z, p)))
                for indice = 1:length(lista(i, j, z, p))
                    if(ischar(lista(i, j, z, p){indice}))
                        materiales = materiales + 1;
                        for r = 1:length(dat_comp{1})
                            if(strcmp(lista(i, j, z, p){indice}, dat_comp{1}{r}))%Es un acierto
                                materiales_acertados = materiales_acertados + 1;
                                dat_comp{1}{r} = [];
                                break;
                            end
                        end
                    end
                end
            end
            Primer_indicador(i, j, z, p) = materiales_acertados;

            %Falsos positivos (solo si se han acertado los que habia en txt): Materiales en lista (que no esten en
            %archivo) - Materiales en archivo txt
            Segundo_indicador(i, j, z, p) = materiales - materiales_acertados;
        end
    end
end
```

Figura 28. Bucle para probar los distintos valores de parámetros

Cuando se terminan los bucles se pasa a calcular los indicadores generales.

El primer indicador general es un porcentaje de los materiales acertados respecto de todos los materiales que se han procesado.

El segundo indicador general un porcentaje del número de falsos positivos que se han ido obteniendo respecto los materiales acertados que ha habido.

```
Primer_general = sum(Primer_indicador, 4)*100/materiales_totales;
Segundo_general = sum(Segundo_indicador, 4)*100./(sum(Segundo_indicador, 4) + sum(Primer_indicador, 4));
```

Figura 29. Indicadores generales

Después de calcular los 2 indicadores que ayudarán a saber cuáles son los valores óptimos, guardamos los datos obtenidos y los abrimos en el último script.

El último script consiste en mostrar los datos de las 2 matrices (porcentaje de aciertos y número de falsos positivos) como una imagen que utiliza un rango de colores usando el comando "imagesc".

Con estas imágenes podemos investigar qué combinación es la mejor.



```
load('20200905_pos_3_25_int_0.1_0.9_pesoInt_0.8.mat')
%%% MOSTRAMOS LAS GRÁFICAS %%%

for z = 1:numel(peso_posicion)
    figure,
    [X,Y] = meshgrid(tolerancia_intensidad,tolerancia_posicion);
    surf(X, Y, Primer_general(:,z));hold on
    imagesc(tolerancia_posicion, tolerancia_intensidad, Primer_general(:,z))
    title(['Positive detections (%). Weights [P-B-I]: [' num2str(peso_posicion(z)) '-' num2str(peso_bandas(z)) '-' num2str(peso_intensidad(z)) ']]')
    ylabel('Intensity tolerance')
    xlabel('Position tolerance')
    colorbar
    set(gcf,'Position',[100 200 600 450])

    figure,
    imagesc(tolerancia_posicion, tolerancia_intensidad, Segundo_general(:,z))
    title(['False positives (#). Weights [P-B-I]: [' num2str(peso_posicion(z)) '-' num2str(peso_bandas(z)) '-' num2str(peso_intensidad(z)) ']]')
    ylabel('Intensity tolerance')
    xlabel('Position tolerance')
    colorbar
    set(gcf,'Position',[700 200 600 450])
end
```

Figura 30. Código para mostrar gráficas

### 5.3 Anexo 3: Gráficas de espectros puros

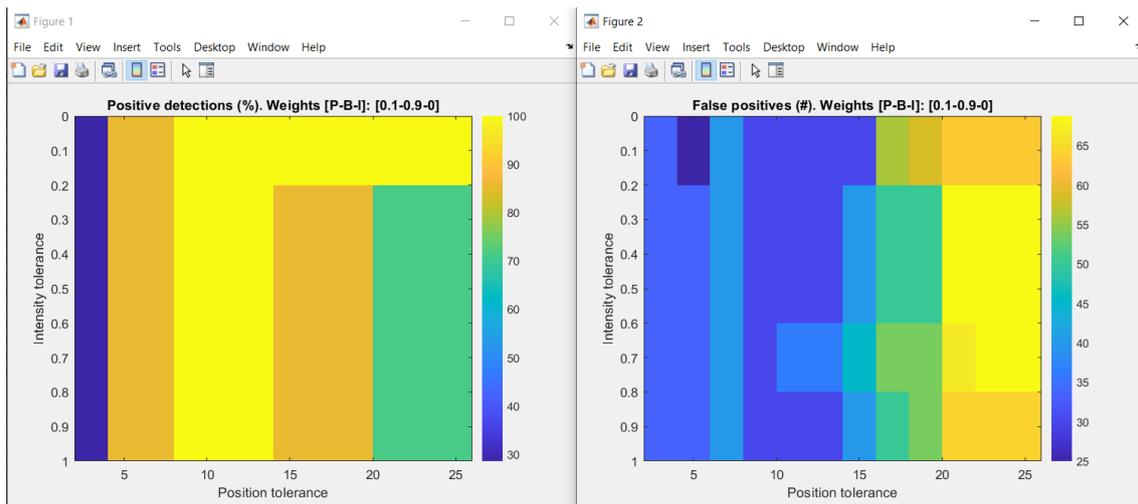


Figura 31. Combinación [P-B-I] = [0.1-0.9-0]

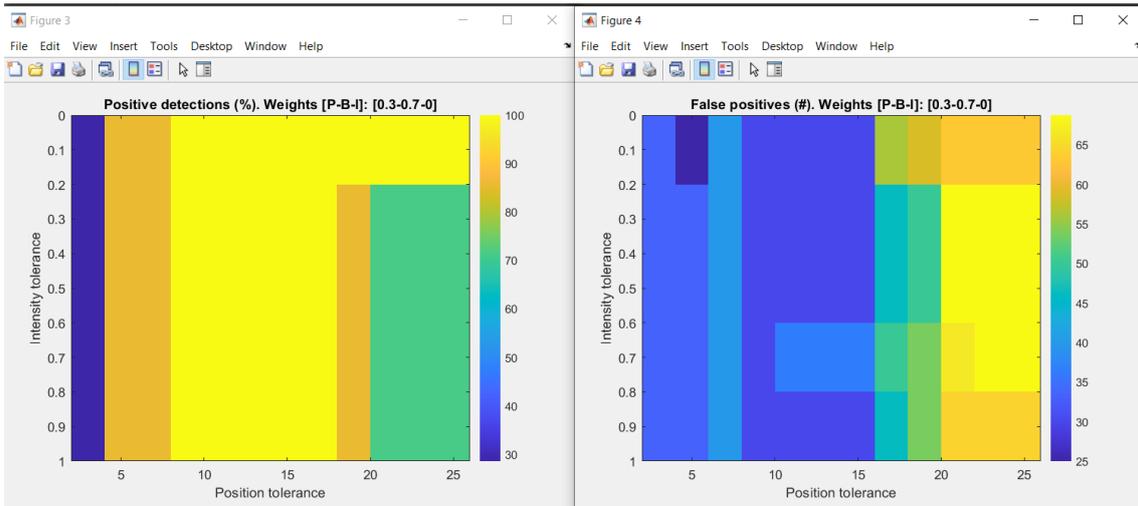


Figura 32. [P-B-I] = [0.3-0.7-0]

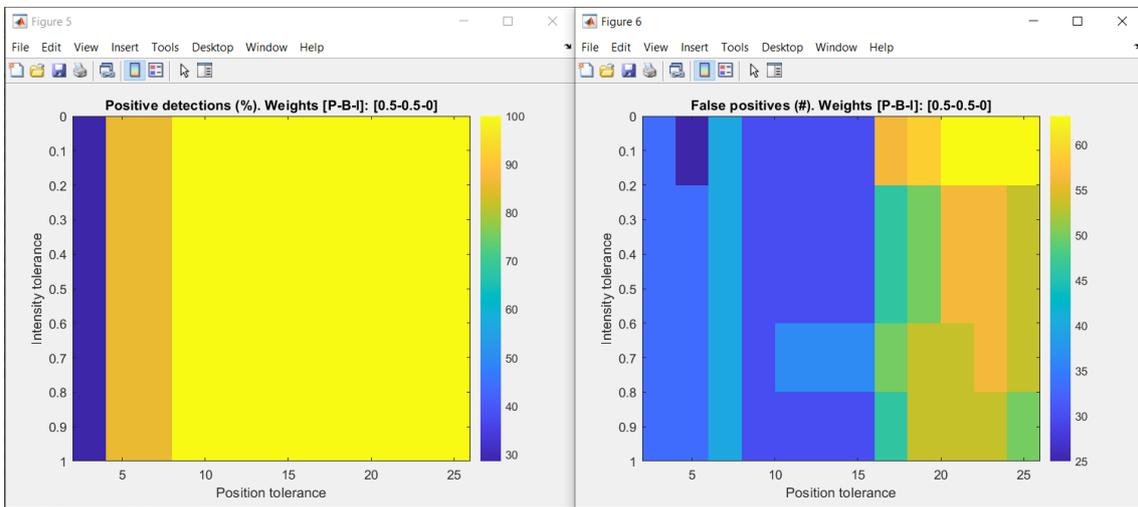


Figura 33. [P-B-I] = [0.5-0.5-0]

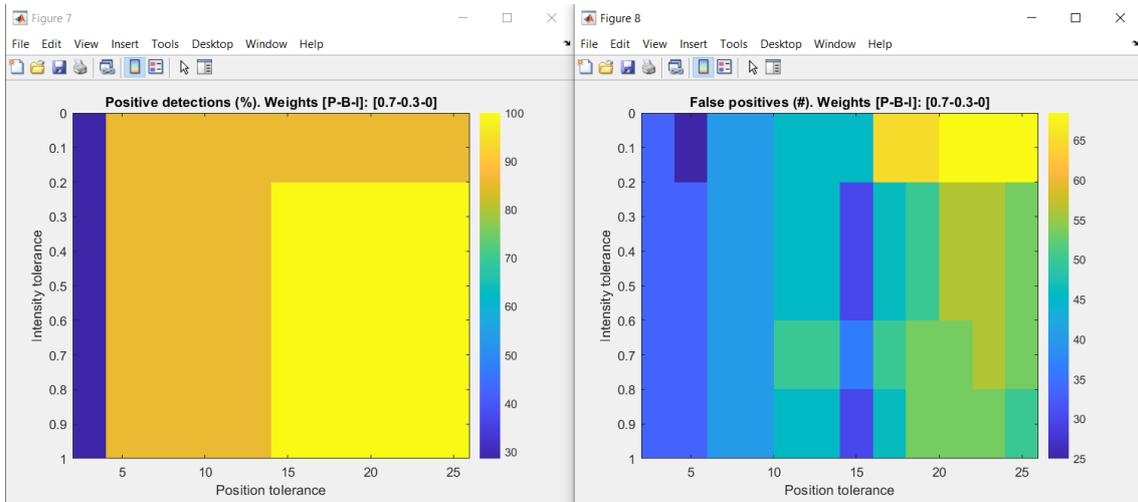


Figura 34. [P-B-I] = [0.7-0.3-0]

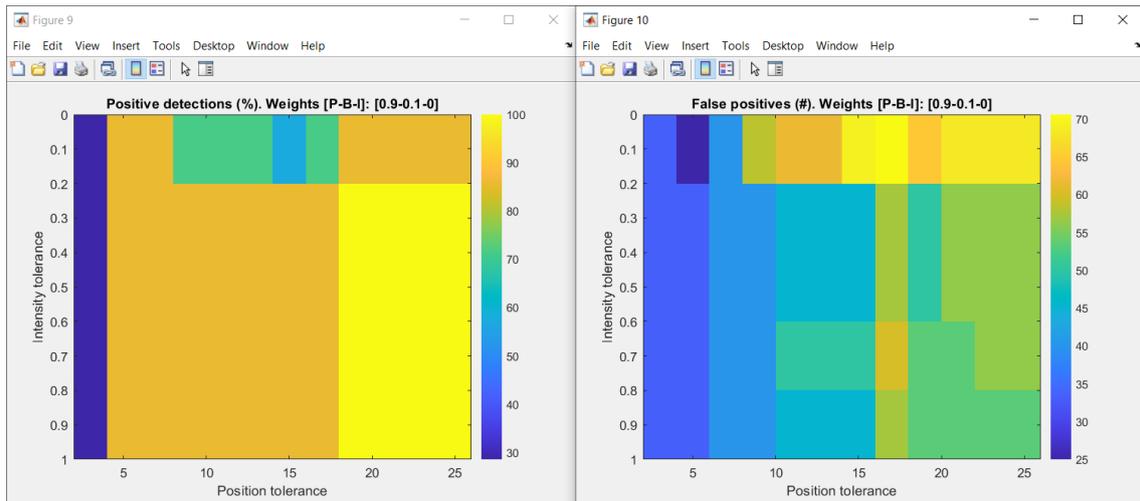


Figura 35.  $[P-B-I] = [0.9-0.1-0]$

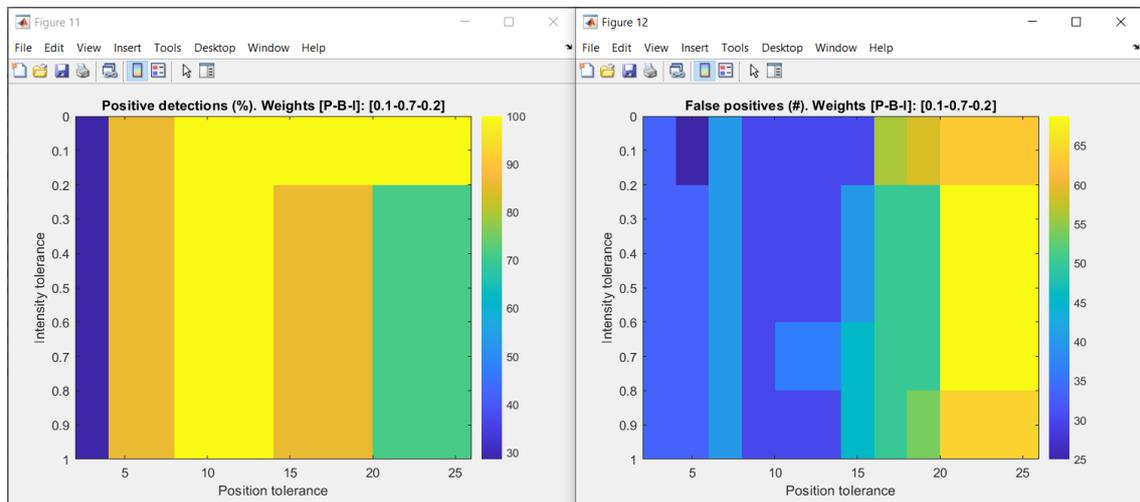


Figura 36.  $[P-B-I] = [0.1-0.7-0.2]$

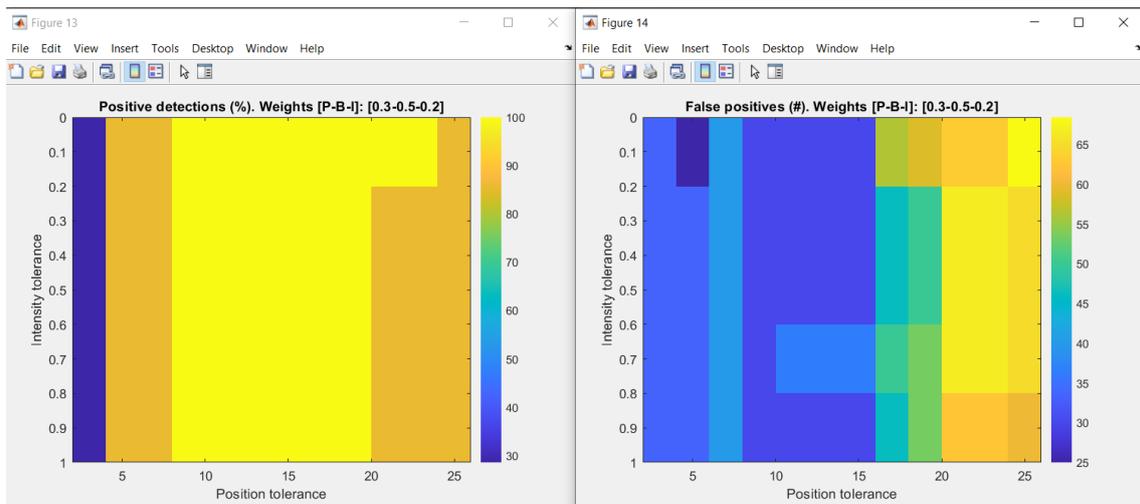


Figura 37.  $[P-B-I] = [0.3-0.5-0.2]$

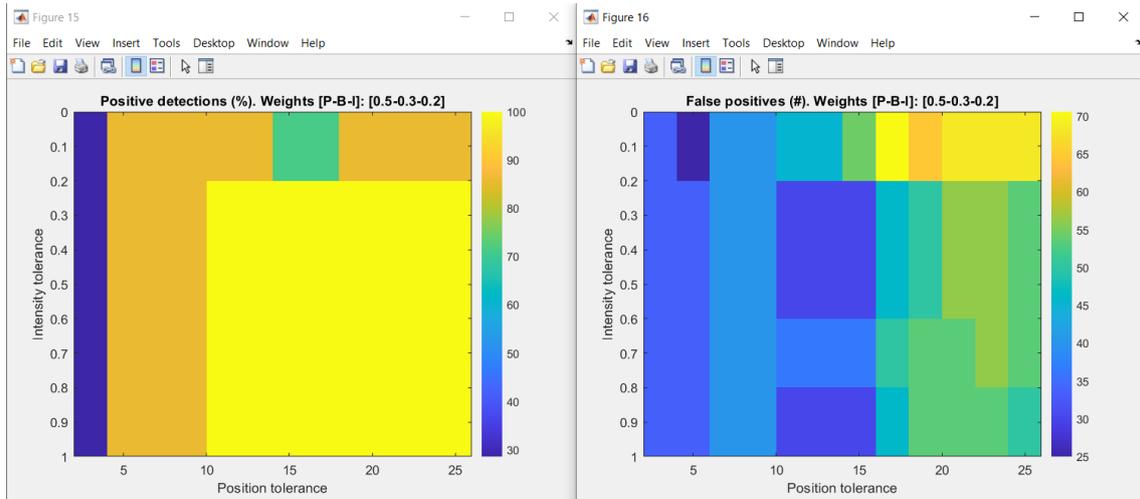


Figura 38.  $[P-B-I] = [0.5-0.3-0.2]$

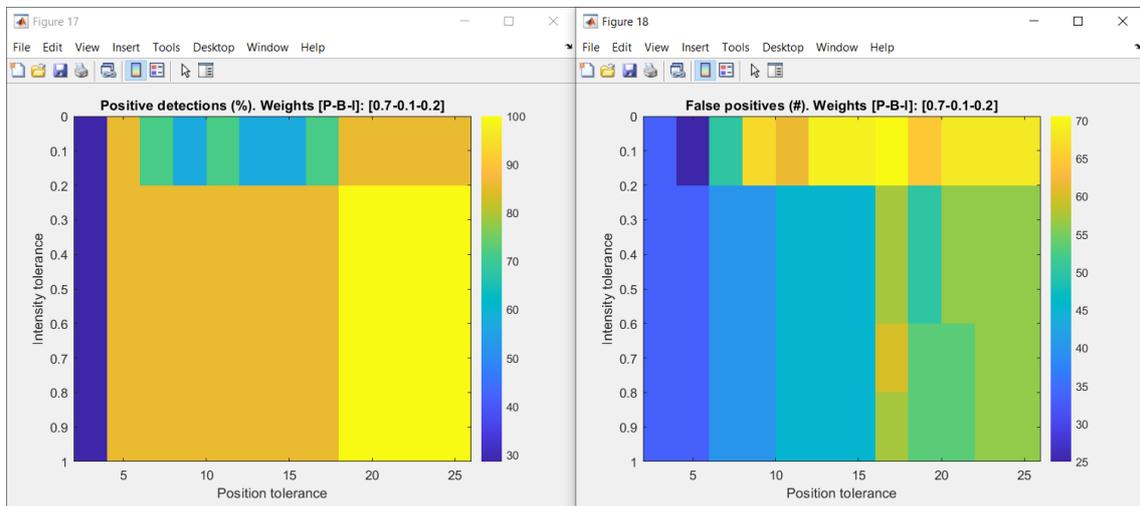


Figura 39.  $[P-B-I] = [0.7-0.1-0.2]$

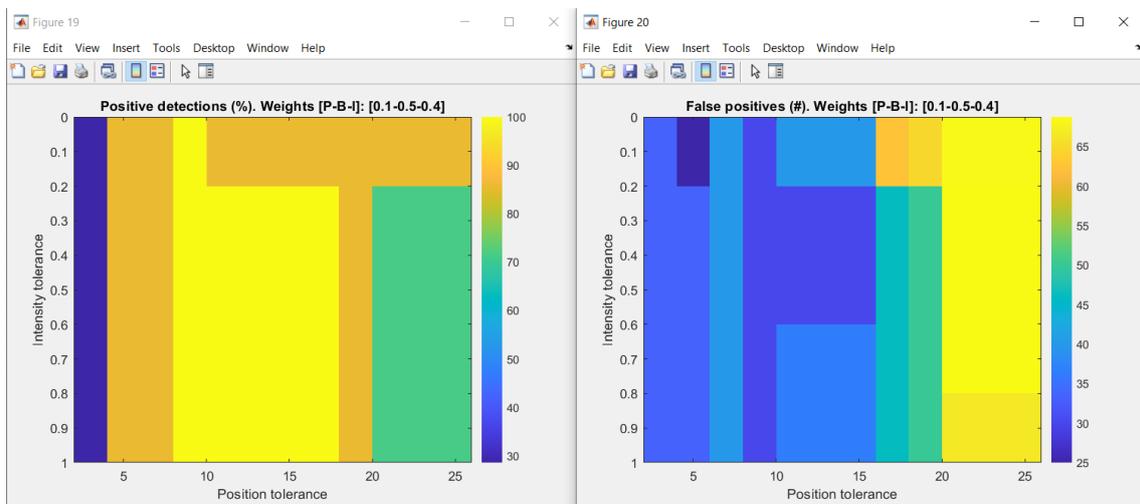


Figura 40.  $[P-B-I] = [0.1-0.5-0.4]$

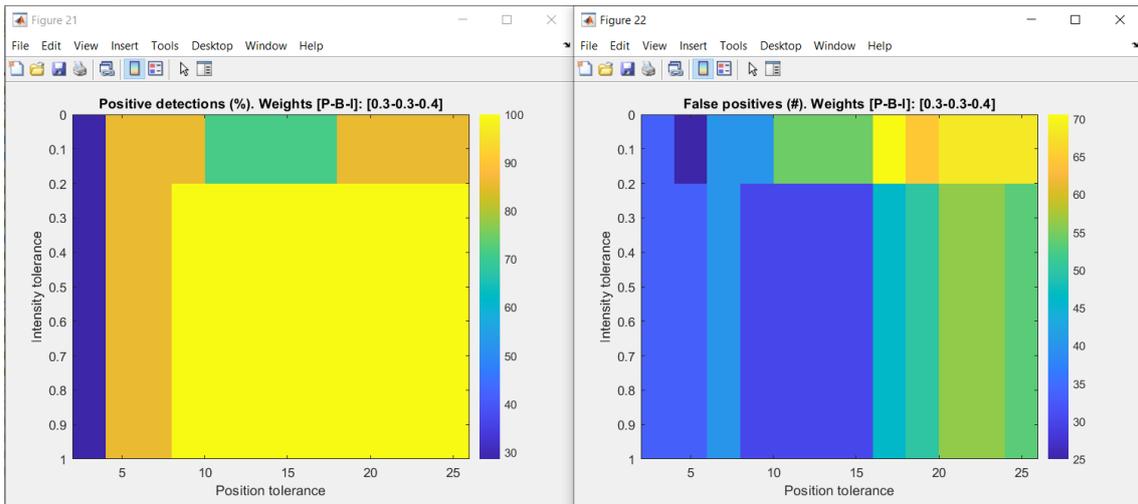


Figura 41. [P-B-I] = [0.3-0.3-0.4]

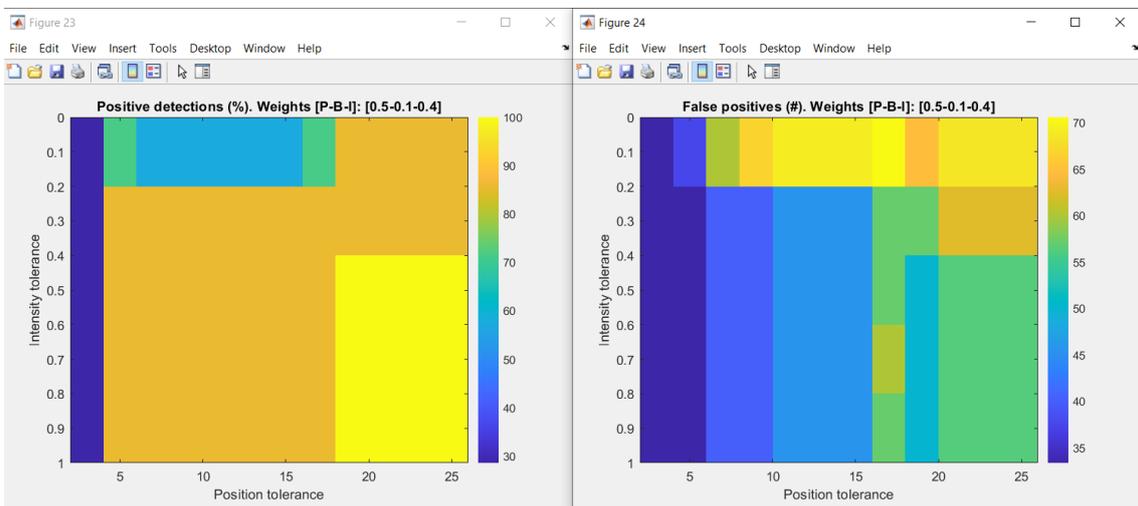


Figura 42. [P-B-I] = [0.5-0.1-0.4]

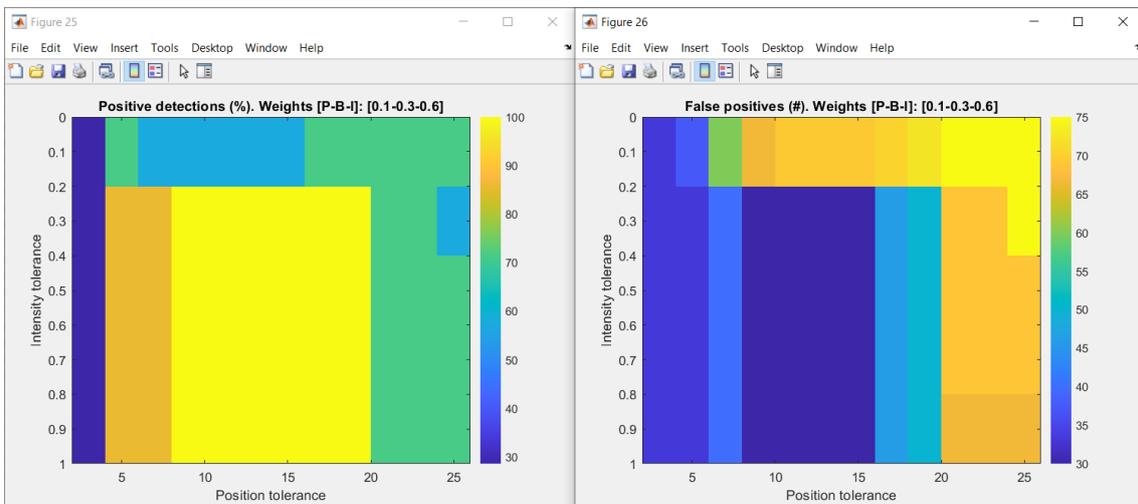


Figura 43. [P-B-I] = [0.1-0.3-0.6]

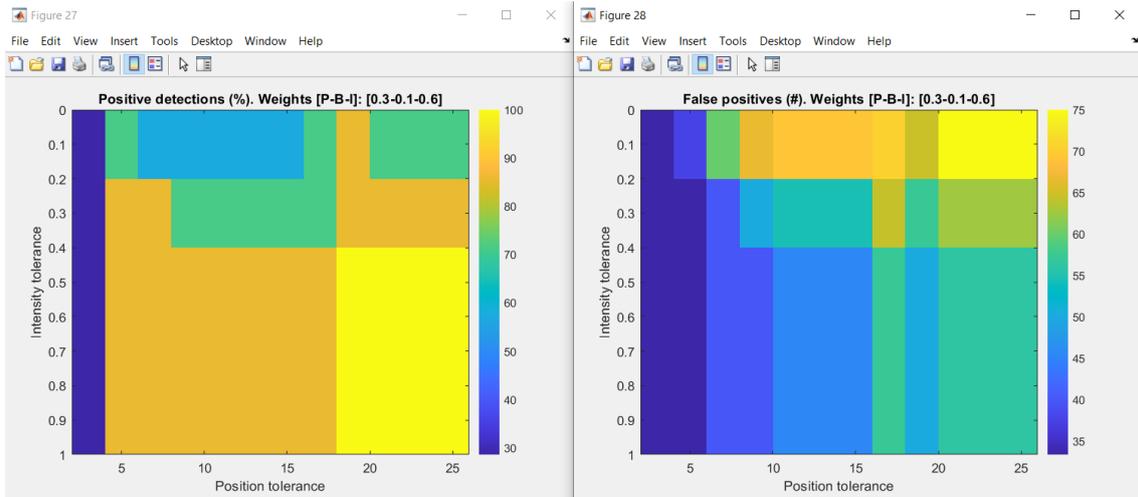


Figura 44. [P-B-I] = [0.3-0.1-0.6]

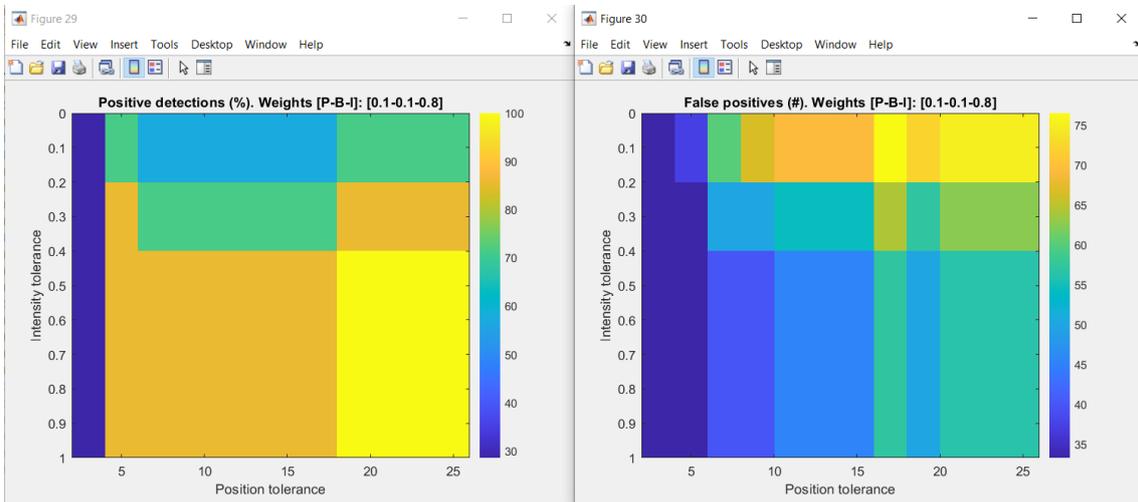


Figura 45. [P-B-I] = [0.1-0.1-0.8]

## 5.4 Anexo 4: Gráficas de espectros mezcla

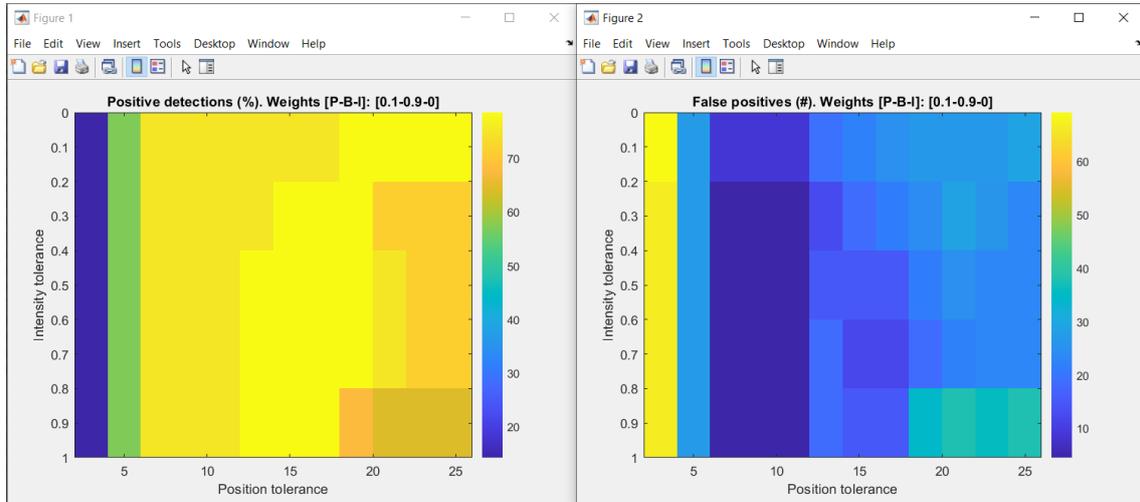


Figura 46.  $[P-B-I] = [0.1-0.9-0]$

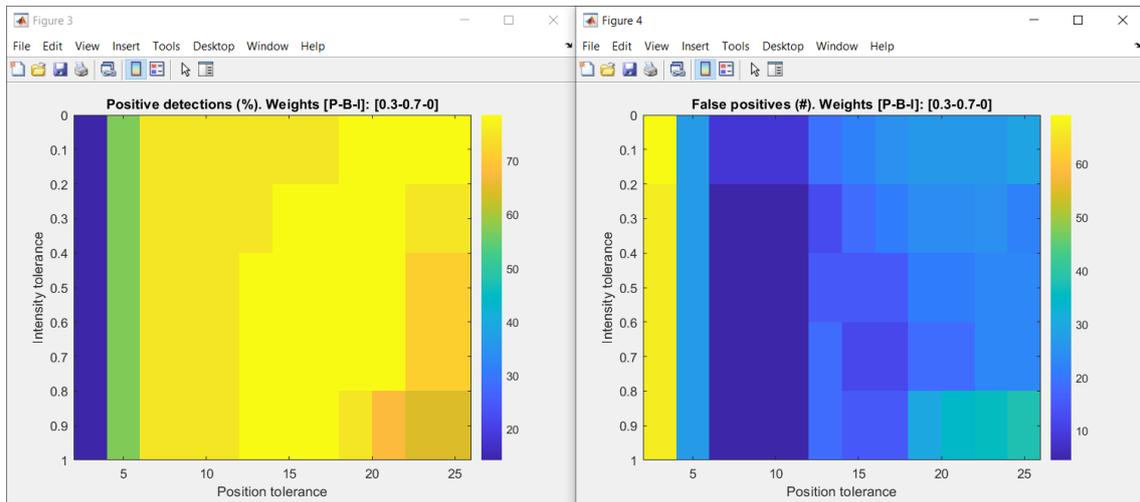


Figura 47.  $[P-B-I] = [0.3-0.7-0]$

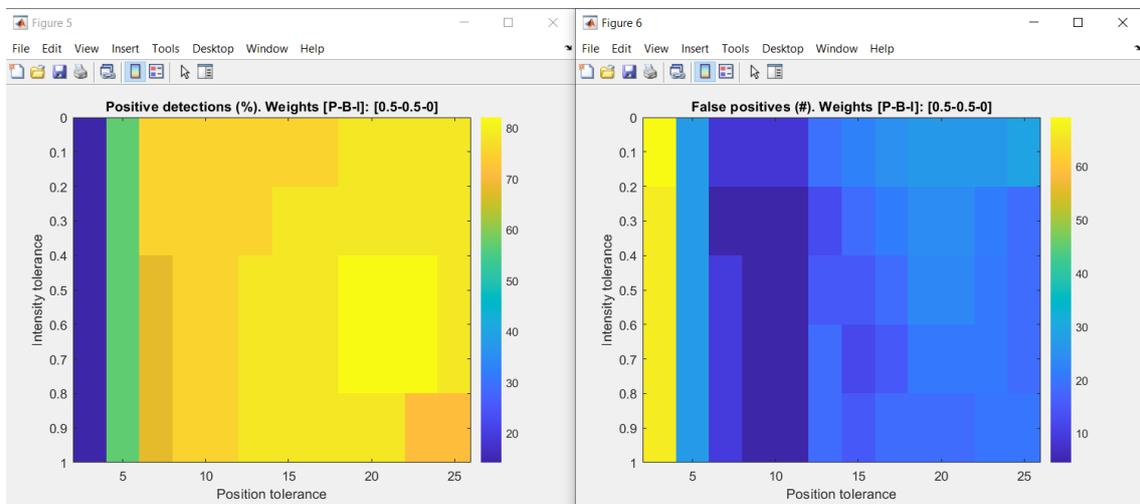


Figura 48.  $[P-B-I] = [0.5-0.5-0]$

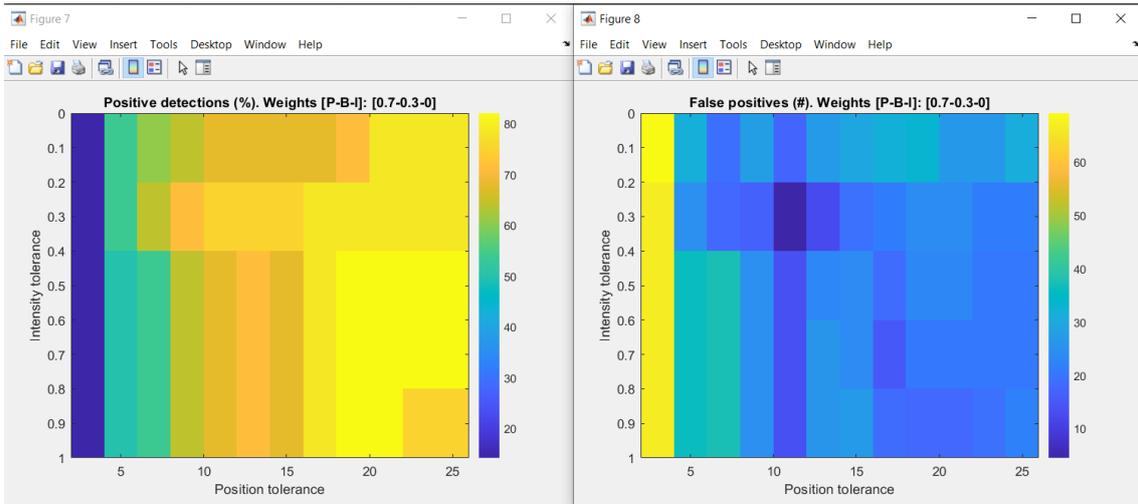


Figura 49. [P-B-I] = [0.7-0.3-0]

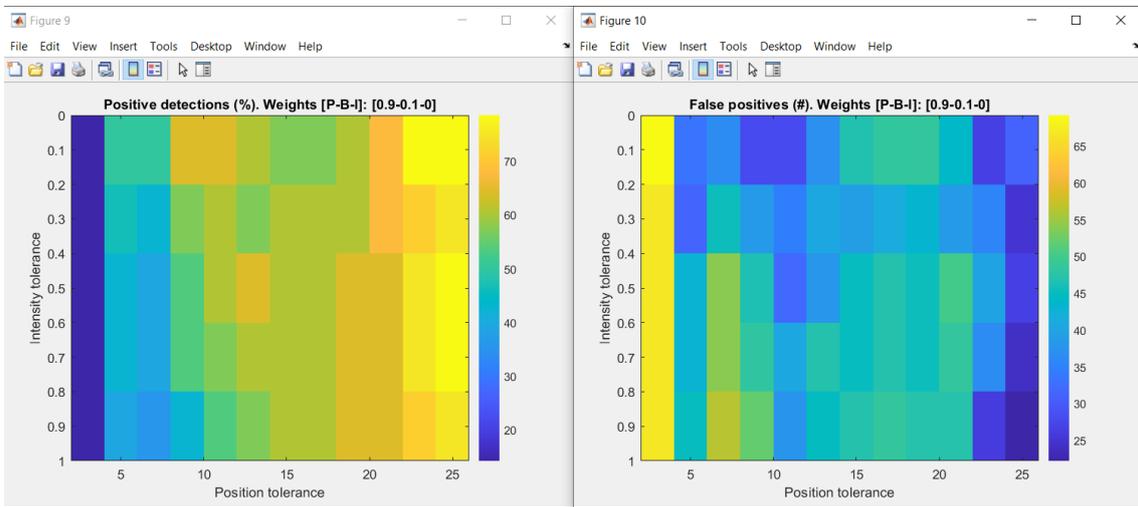


Figura 50. [P-B-I] = [0.9-0.1-0]

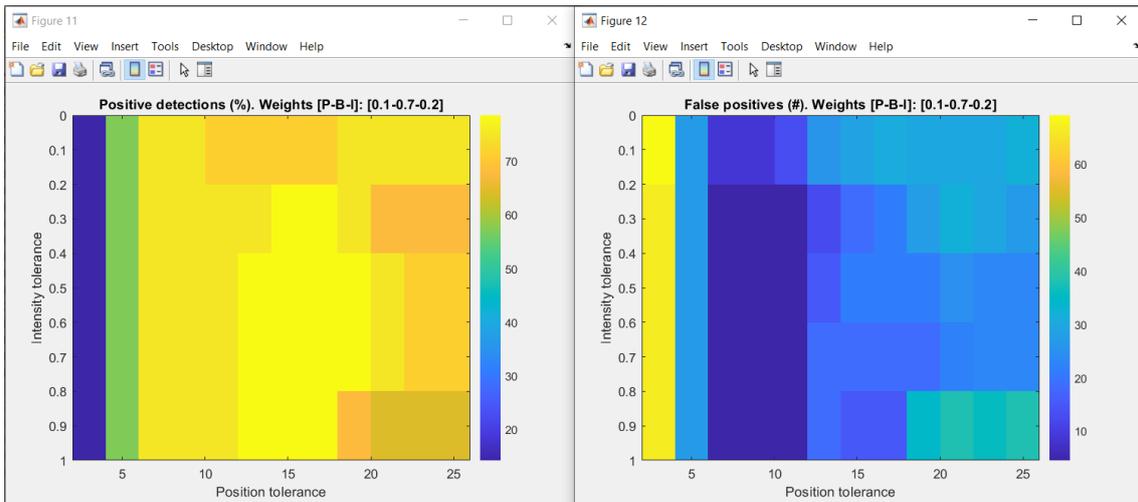


Figura 51. [P-B-I] = [0.1-0.7-0.2]

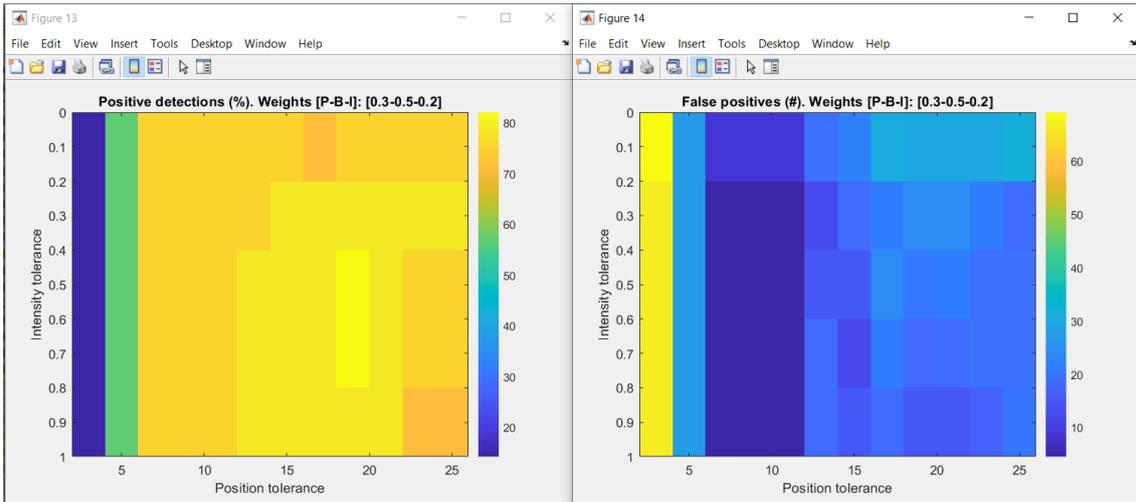


Figura 52. [P-B-I] = [0.3-0.5-0.2]

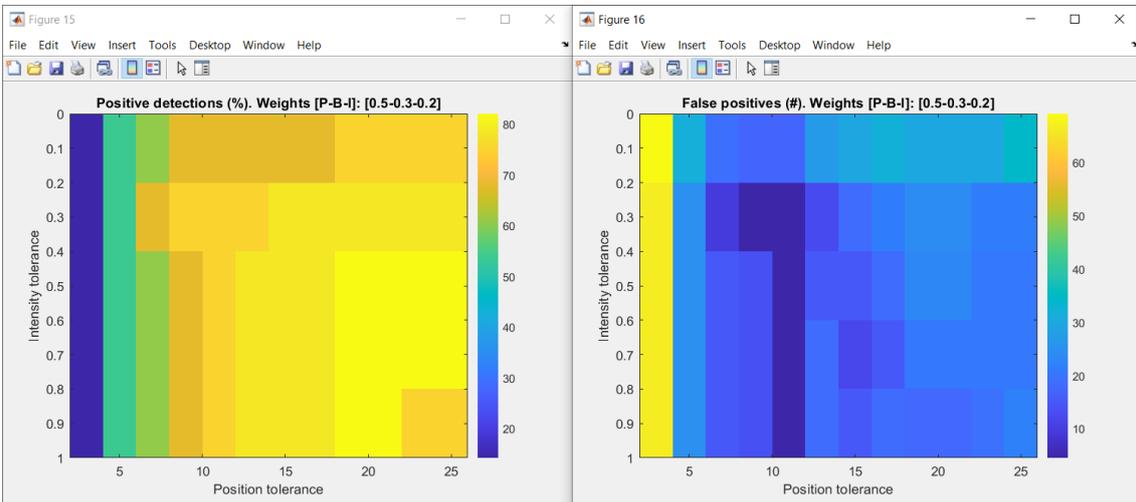


Figura 53. [P-B-I] = [0.5-0.3-0.2]

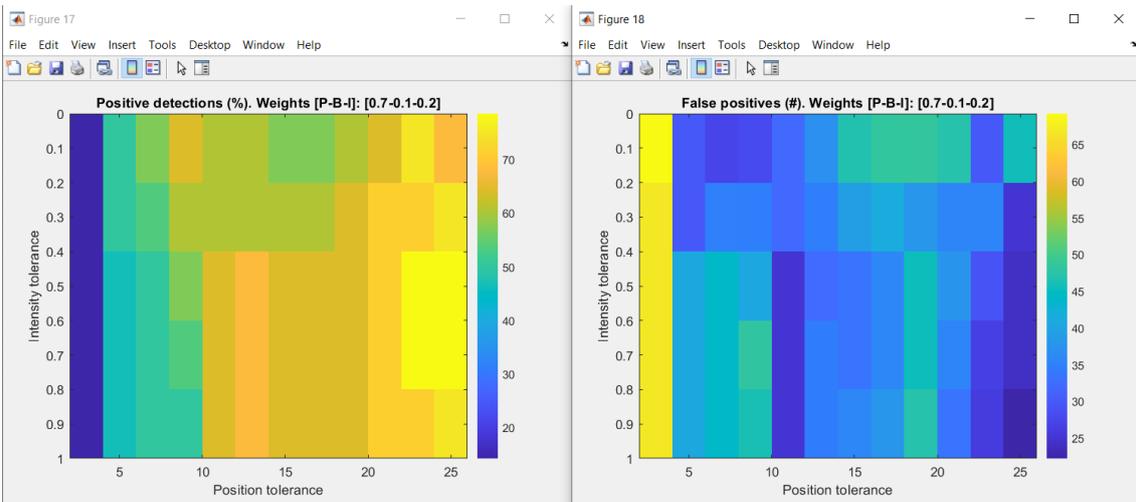


Figura 54. [P-B-I] = [0.7-0.1-0.2]

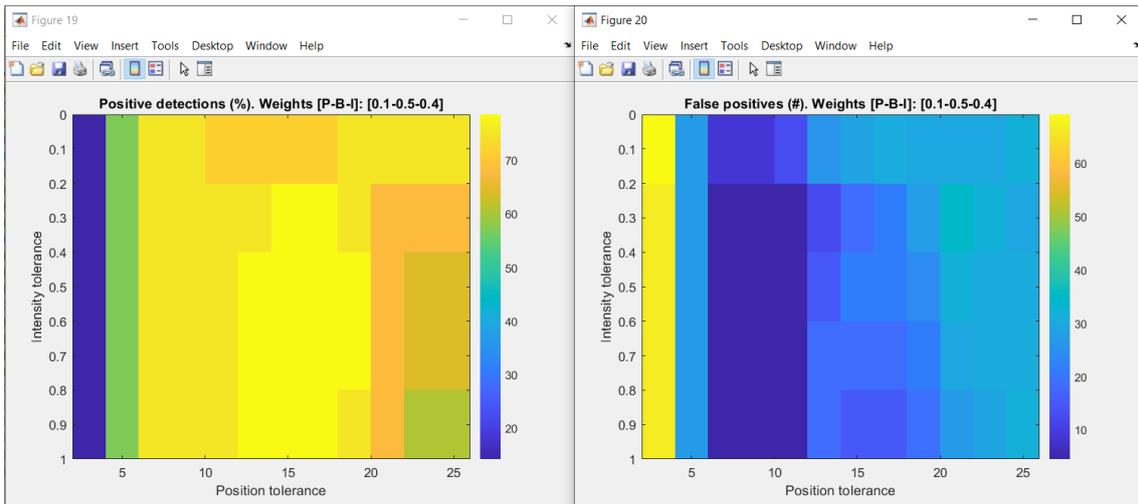


Figura 55. [P-B-I] = [0.1-0.5-0.4]

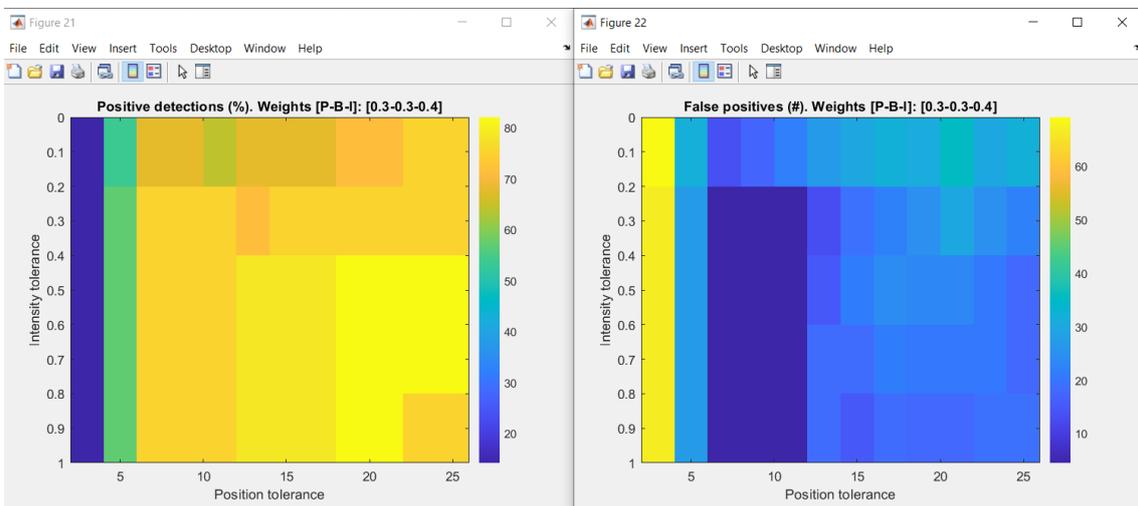


Figura 56. [P-B-I] = [0.3-0.3-0.4]

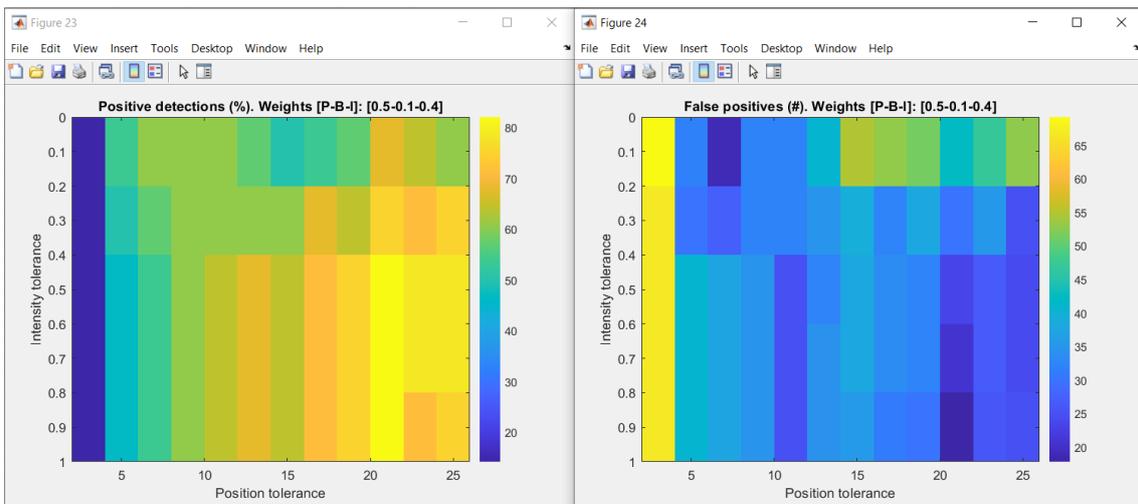


Figura 57. [P-B-I] = [0.5-0.1-0.4]

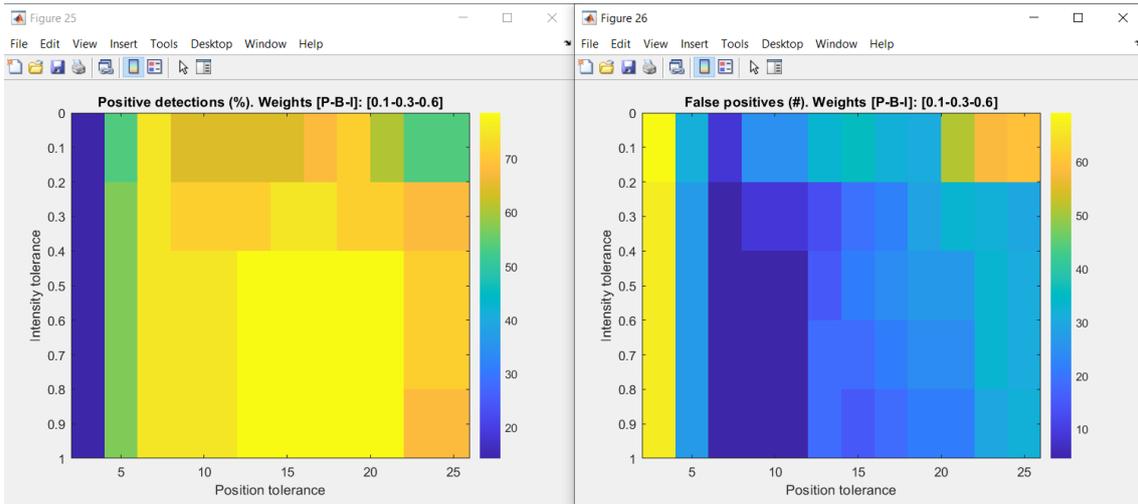


Figura 58.  $[P-B-I] = [0.1-0.3-0.6]$

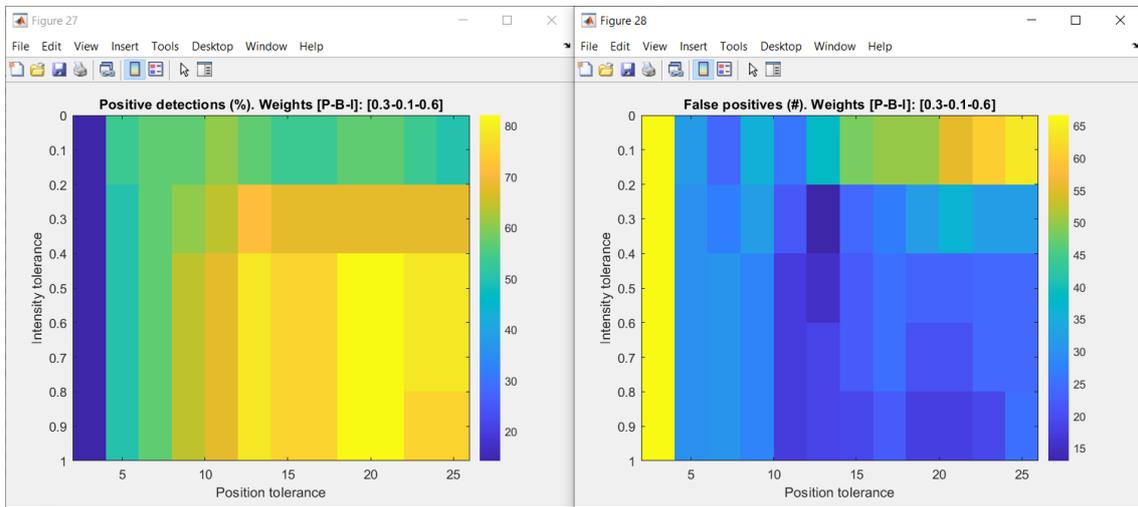


Figura 59.  $[P-B-I] = [0.3-0.1-0.6]$

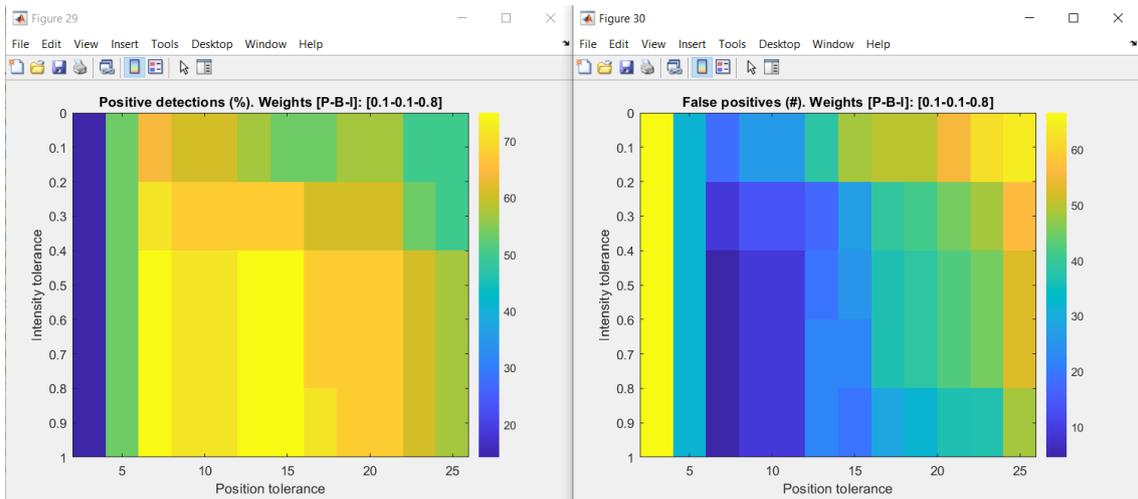


Figura 60.  $[P-B-I] = [0.1-0.1-0.8]$



## 5.5 Anexo 5: Gráficas utilizando todos los espectros problema

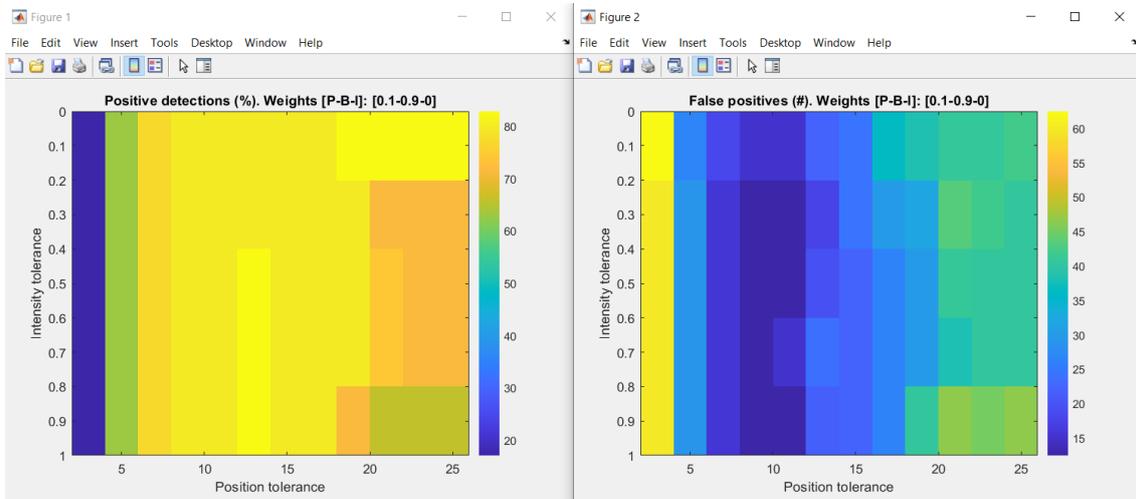


Figura 61. [P-B-I] = [0.1-0.9-0]

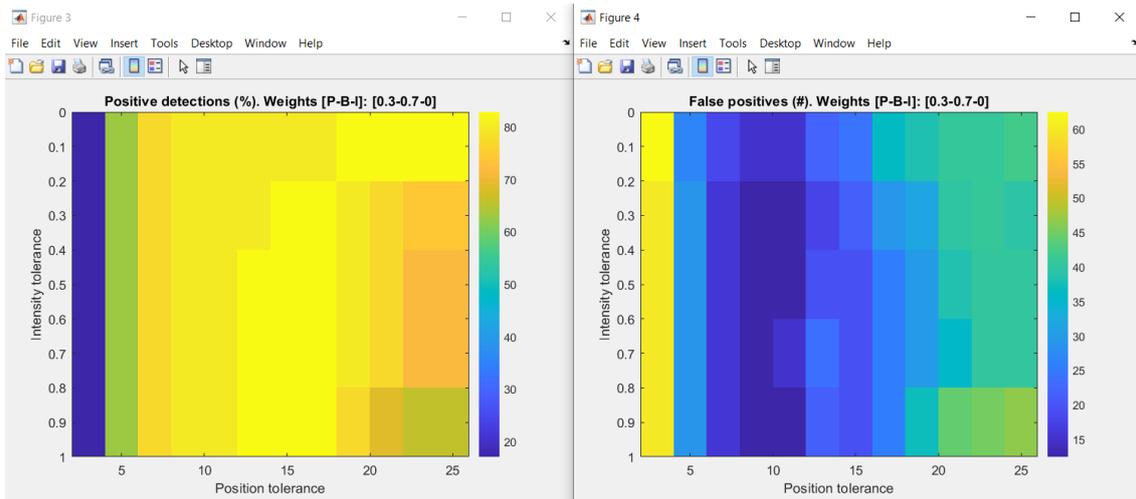


Figura 62. [P-B-I] = [0.3-0.7-0]

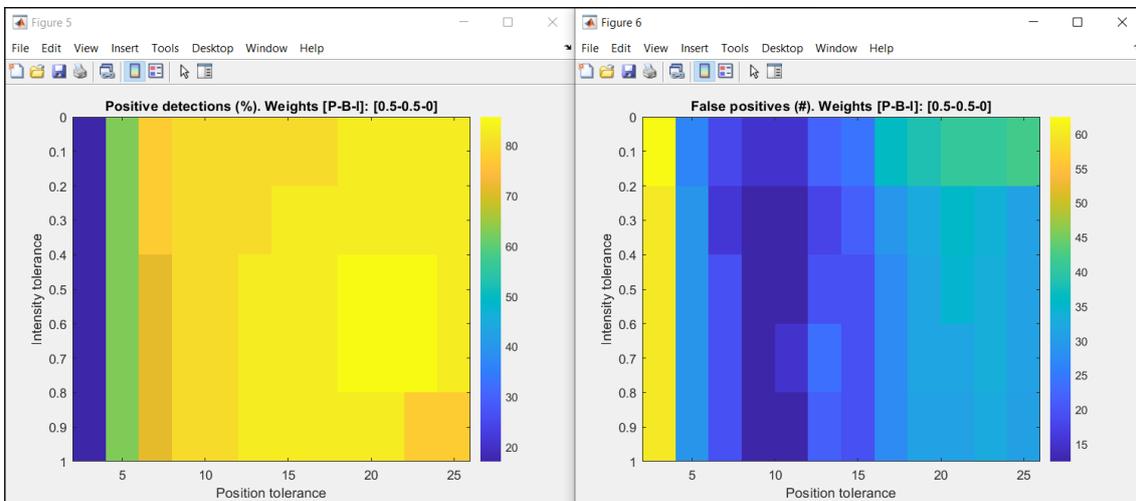


Figura 63. [P-B-I] = [0.5-0.5-0]

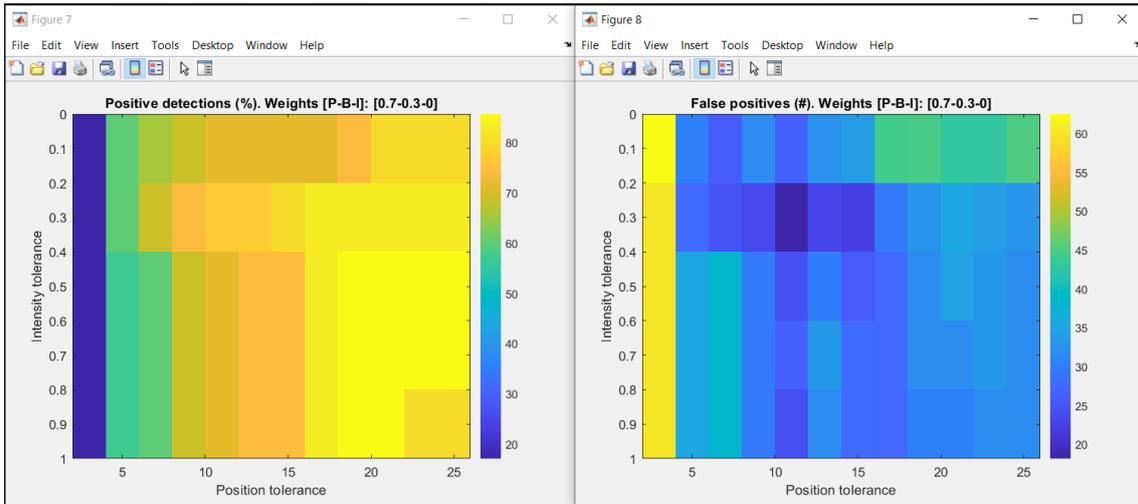


Figura 64. [P-B-I] = [0.7-0.3-0]

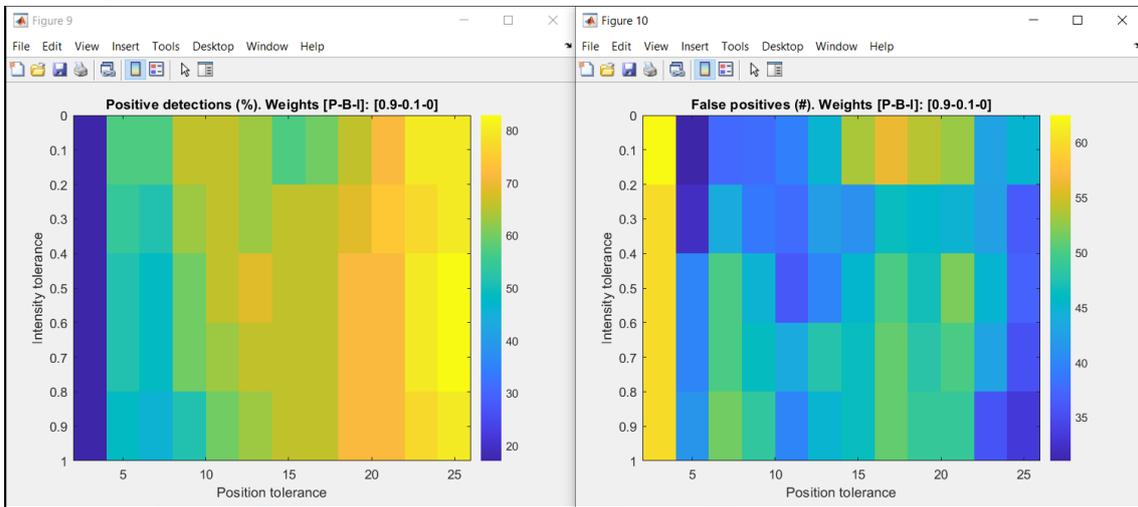


Figura 65. [P-B-I] = [0.9-0.1-0]

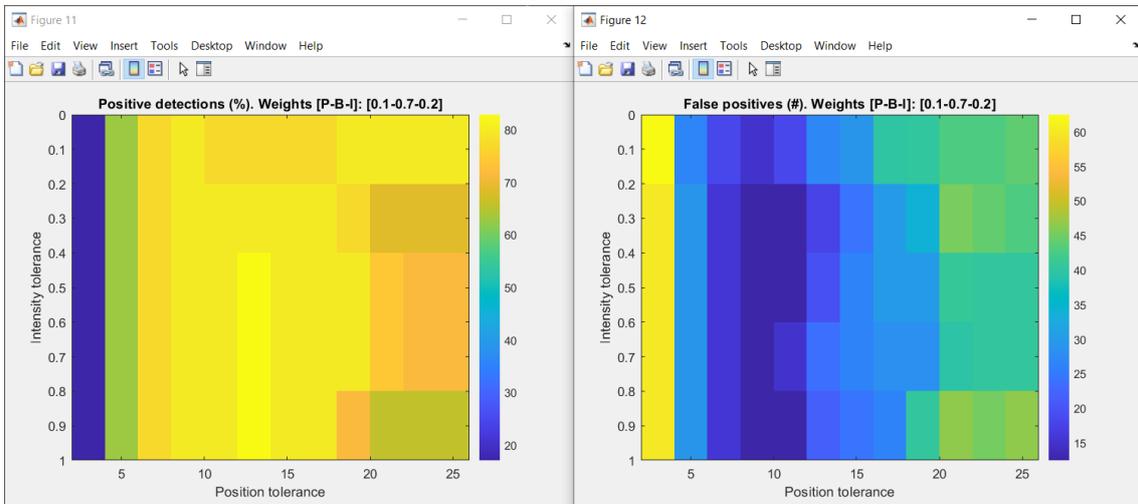


Figura 66. [P-B-I] = [0.1-0.7-0.2]

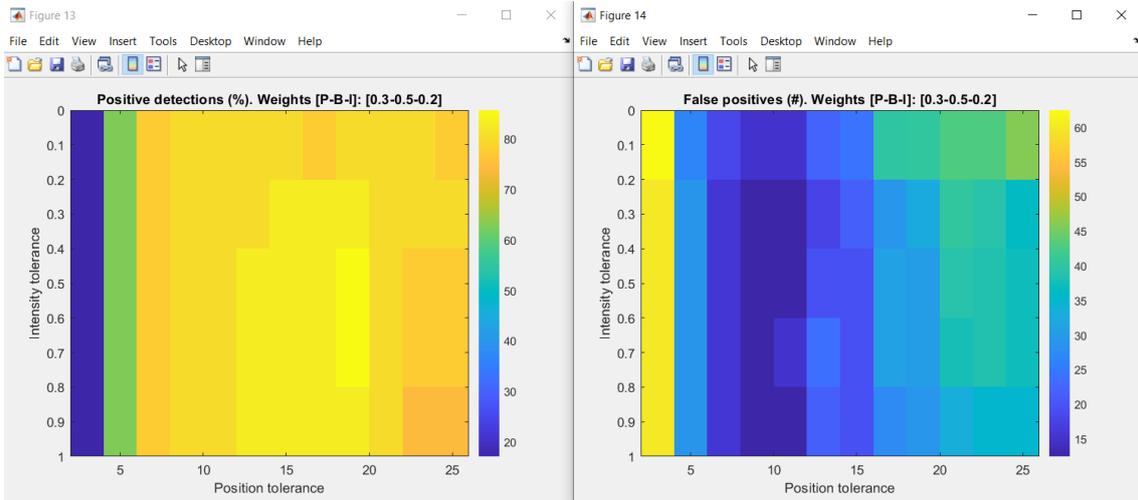


Figura 67. [P-B-I] = [0.3-0.5-0.2]

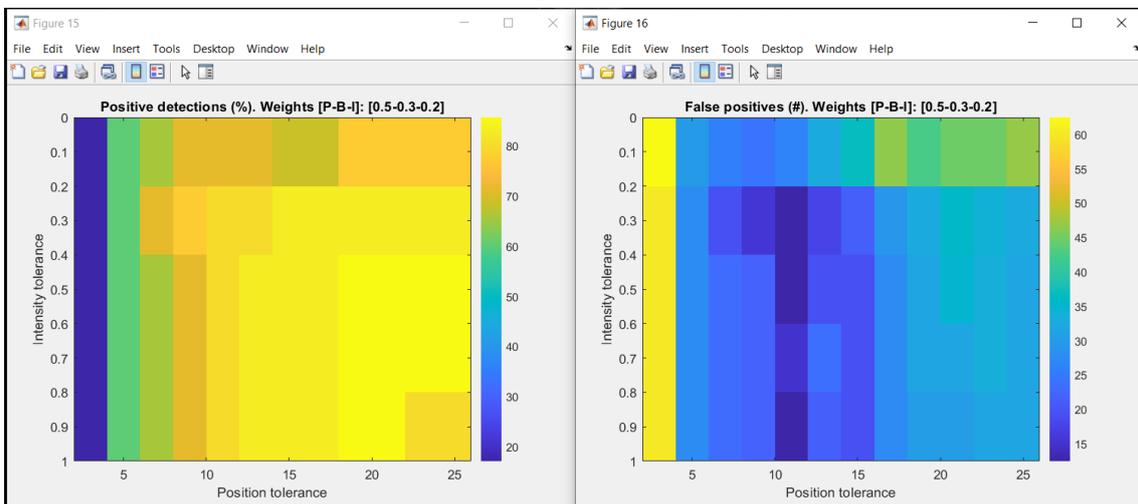


Figura 68. [P-B-I] = [0.5-0.3-0.2]

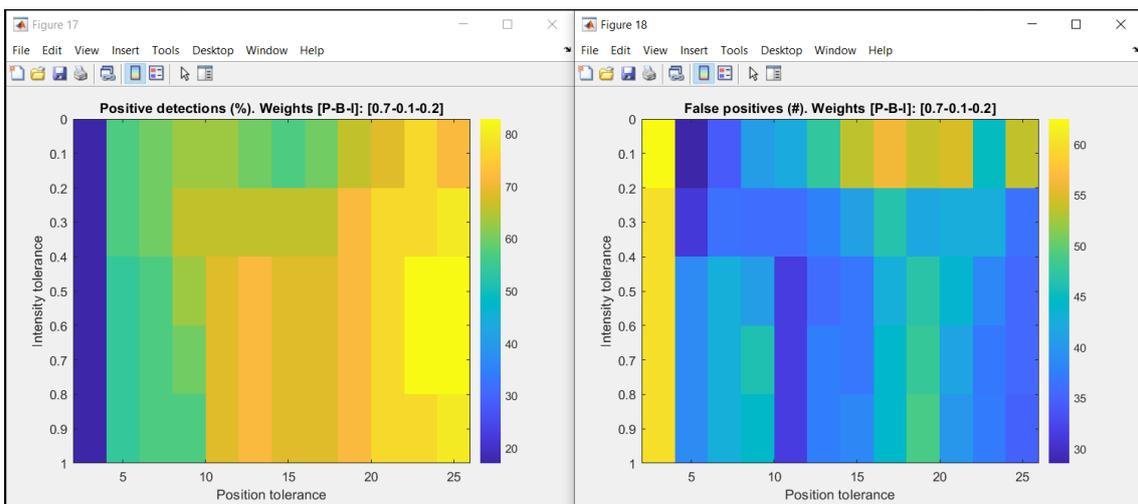


Figura 69. [P-B-I] = [0.7-0.1-0.2]

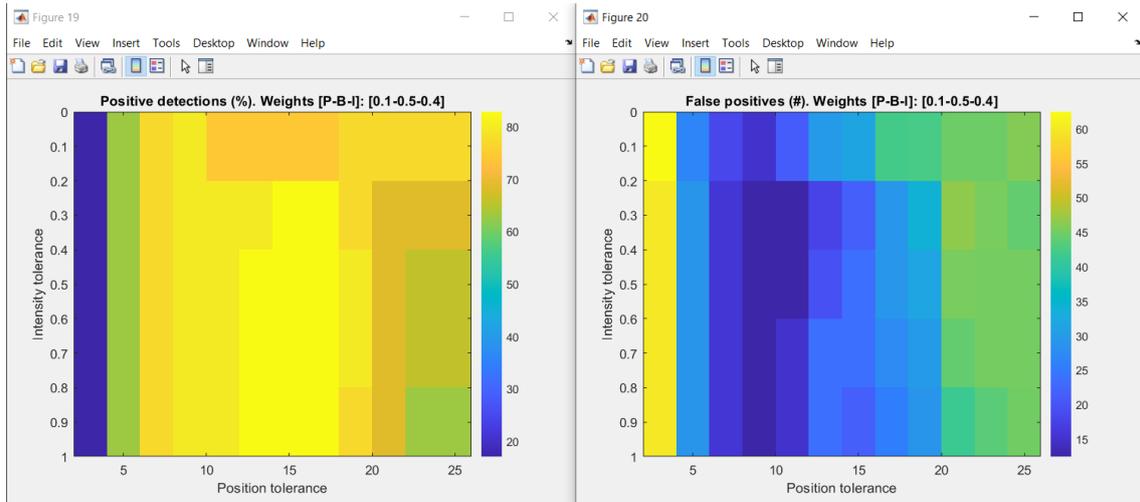


Figura 70. [P-B-I] = [0.1-0.5-0.4]

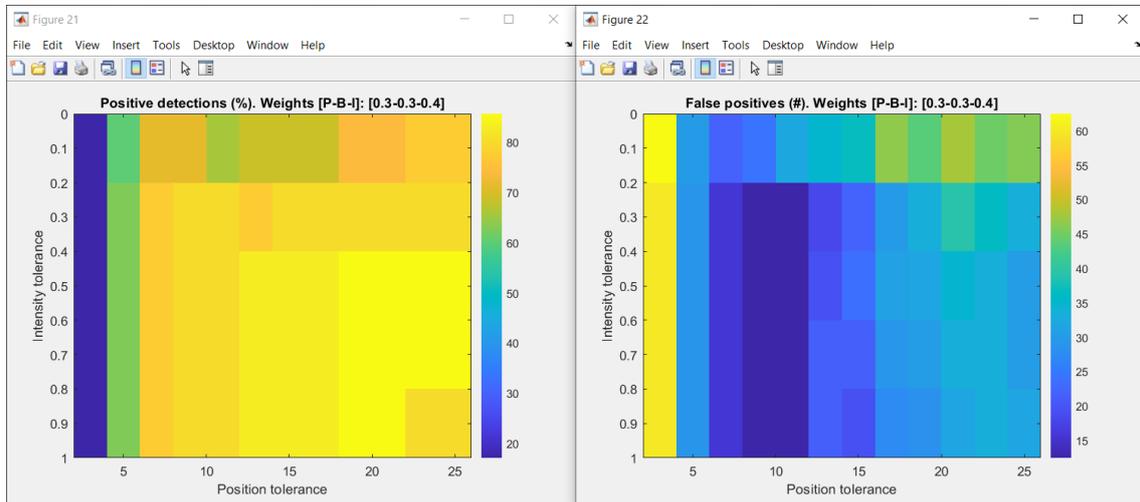


Figura 71. [P-B-I] = [0.3-0.3-0.4]

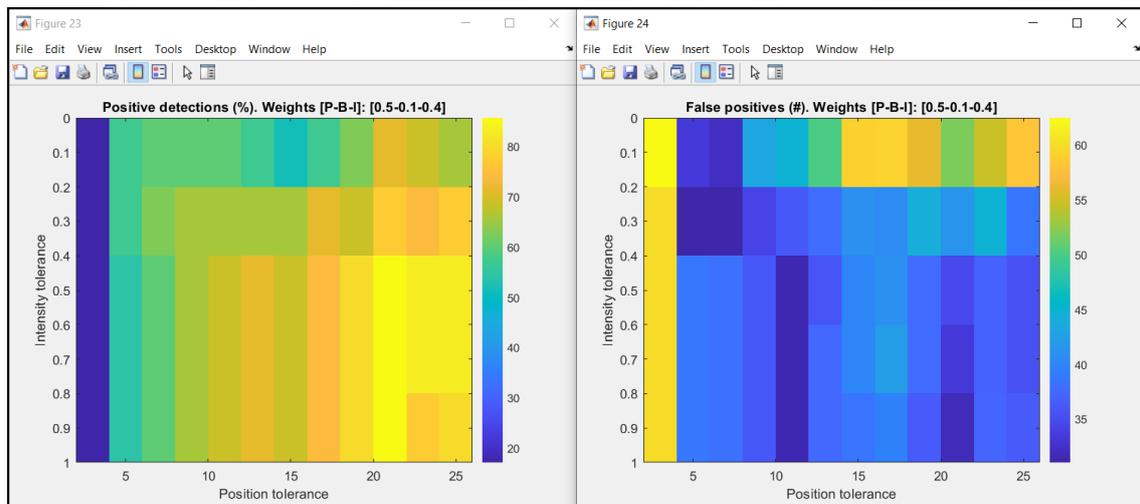


Figura 72. [P-B-I] = [0.5-0.1-0.4]

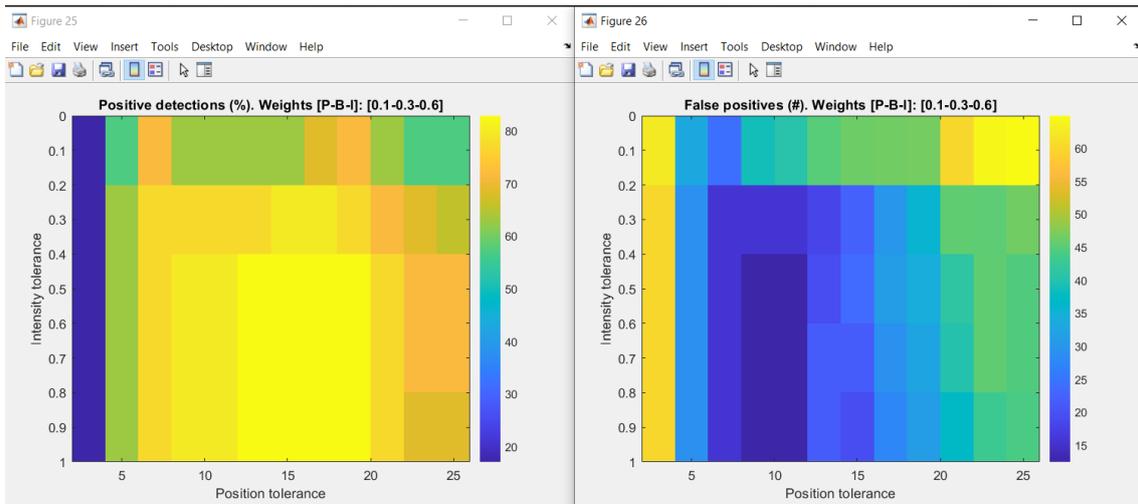


Figura 73. [P-B-I] = [0.1-0.3-0.6]

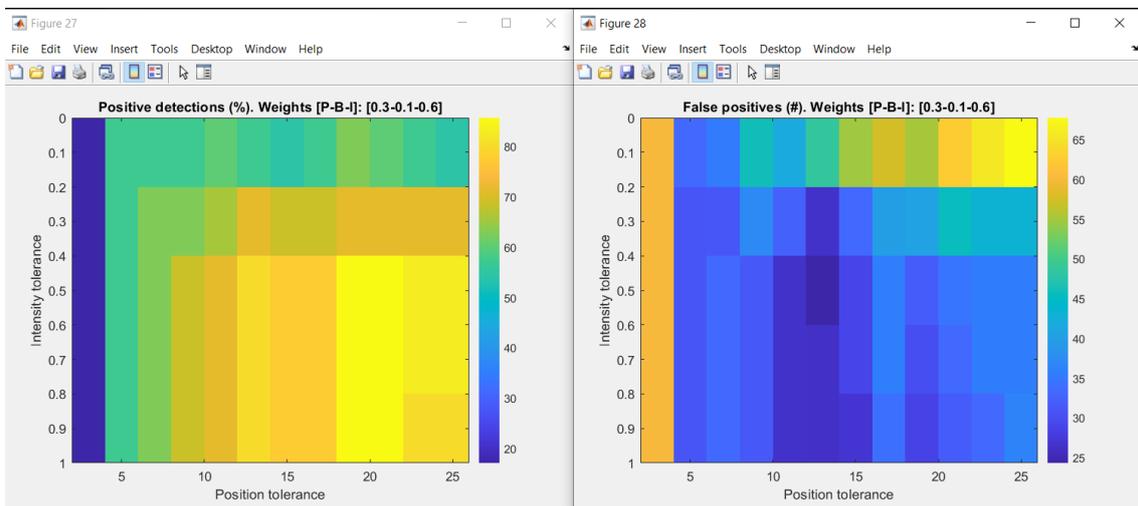


Figura 74. [P-B-I] = [0.3-0.1-0.6]

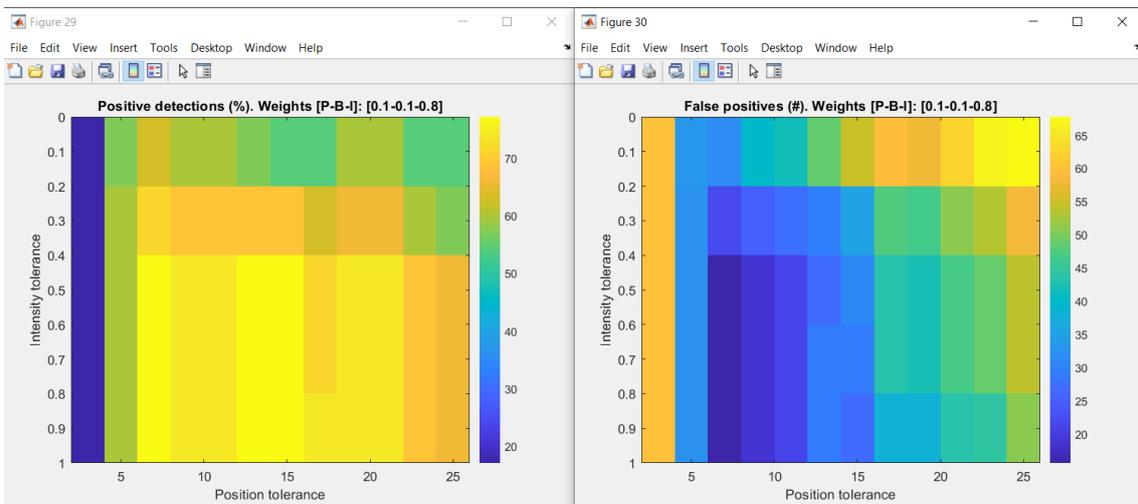


Figura 75. [P-B-I] = [0.1-0.1-0.8]



## 5.6 Anexo 6: Análisis detallado de algunos espectros individuales

Atendiendo al relativamente elevado porcentaje de falsos positivos que surgen durante la parametrización del algoritmo se ha decidido analizar espectros de algunos materiales individuales y espectros de mezclas para ver que ocurre concretamente, intentando dar respuestas y conocer más profundamente el algoritmo a nivel funcional.

Los espectros individuales que se han seleccionado son de los materiales olivino y calcita.

Como espectro mezcla se analizará una mezcla de cuarzo y calcita en una concentración de 75% y 25% respectivamente.

Los parámetros que se han elegido para analizar los espectros son:

- Tolerancia de intensidad = 0,4.
- Tolerancia de posición = 9.
- Peso de intensidad = 0,2.
- Peso de posición = 0,2.
- Peso de bandas = 0,6.

### 5.6.1 Calcita

El espectro problema de la calcita consta de los siguientes picos en el archivo xml.

- Banda 1: intensidad = 0,22. Posición = 156,19.
- Banda 2: intensidad = 0,46. Posición = 285,19.
- Banda 3: intensidad = 0,13. Posición = 715,19.
- Banda 4: intensidad = 1. Posición = 1090,19.

El código que se ha utilizado está descrito a partir del punto 5.1.

Cuando leemos los posibles materiales que cumplen con que su pico principal (valor de intensidad igual a 1) se encuentra en el rango de posiciones 1081,19-1099,19 nos sale una lista con los siguientes materiales.

- Siderita. La posición de su pico principal se encuentra en 1083,50.
- Rodocrosita. La posición de su pico principal se encuentra en 1085.
- Magnesita. La posición de su pico principal se encuentra en 1094.
- Epidota. La posición de su pico principal se encuentra en 1085.
- Calcita. La posición de su pico principal se encuentra en 1086,50.
- Aragonita. La posición de su pico principal se encuentra en 1085,50.

En contra de lo que esperábamos, el pico principal del espectro de la calcita en la base de datos difiere del pico principal del espectro problema. Si hubiéramos dado una tolerancia de posición más pequeña, por ejemplo, de 4, nos habrían salido unos resultados falsos.



Con esto sacamos que la tolerancia de posición que tenemos que elegir tiene que ser de un valor ni muy pequeño, como en este caso de 4, ni muy grande porque cogéramos materiales de más.

A continuación, procedemos a ordenar esta lista según el número de bandas que tienen en común, la diferencia de posiciones que tengan sus picos y la diferencia de intensidades que tengan éstos.

Después de la comparación de bandas de los espectros de la base de datos con las bandas del espectro problema se devuelven los siguientes resultados:

nombre	Nº bandas	Bandas	Dif. Intensidad	Dif. Posición
Siderite	5	2	0,02	0,6044
Rodocrosita	6	3	0,0667	0,4189
Magnesita	5	1	0	0,4233
Epidota	8	1	0	0,5767
Calcita	6	4	0,0125	0,2711
Aragonita	4	1	0,5	0,7606

Tabla 5. Resultados de las comparaciones de las bandas

La primera columna indica el número de bandas que tiene cada material según los espectros que se encuentran en la base de datos.

La segunda columna hace referencia a la cantidad de bandas que coinciden con las bandas con el espectro problema.

La tercera columna es la media de la diferencia de intensidades entre los picos de las bandas.

La cuarta columna es la media de la diferencia de posiciones entre los picos de las bandas.

Se puede ver que la calcita es el material que más bandas parecidas tiene con el del espectro problema, un total de 4 bandas. Las intensidades de los picos de la rodocrosita y la magnesita se parecen más a los del espectro problema, pero, en este caso nos sale que la diferencia de intensidades es 0 debido a que la única banda que tienen en común es la del pico principal. Esto se identifica como un posible problema del algoritmo, ya que es un contra-sentido que, por el hecho de tener una única banda coincidente, el indicador sea más favorable a ese material. Por tanto, debería considerarse como **línea futura el incluir el número de bandas detectadas de forma implícita en el cálculo del diferencial en posición e intensidad**. Por ejemplo, dando por defecto un valor 1 a todos los indicadores de posición e intensidad que se usan para calcular Dif. Intensidad y Dif.



Posición, teniendo en cuenta todas las bandas del espectro problema y no solamente las bandas coincidentes.

En este caso, la calcita es el material que más se parece al del espectro problema teniendo la menor diferencia de intensidades (si no contamos la rodocrosita y la magnesita) y la menor diferencia entre posiciones.

Al multiplicar por los pesos, el resultado nos queda de esta manera:

nombre	Dif. Intensidad	Dif. Posición	Bandas	Coincidencia
Siderite	0,0040	0,1209	0,300	57,511
Rodocrosita	0,0133	0,0838	0,1500	75,2889
Magnesita	0	0,08455	0,4500	46,533
Epidota	0	0,1153	0,4500	43,4667
Calcita	0,0025	0,0542	0	94,3278
Aragonita	0,100	0,1521	0,4500	29,7889

*Tabla 6. Resultado final*

La columna “Coincidencia” de la tabla 6 resulta restarle a 1 la suma de los valores de las otras 3 columnas y multiplicar por 100.

La calcita es el material que más se parece al del espectro problema, con un 94,32% por lo que es el material que nos saldría en primer puesto al final de la ejecución del algoritmo.

En este caso no sale ningún falso positivo porque las 4 bandas que tenía el espectro problema han sido usadas para su detección. Por lo que el algoritmo, después de eliminar las bandas que han sido utilizadas, no se volverá a ejecutar.

### **5.6.2 Olivino**

El espectro problema de este material consta de 12 bandas:

- Banda 1: Intensidad = 0,09. Posición = 340,60.
- Banda 2: Intensidad = 0,10. Posición = 485,94.
- Banda 3: Intensidad = 0,13. Posición = 549,07.
- Banda 4: Intensidad = 0,17. Posición = 593,69.
- Banda 5: Intensidad = 0,09. Posición = 639,61.
- Banda 6: Intensidad = 0,87. Posición = 826,69.
- Banda 7: Intensidad = 1. Posición = 858,69.
- Banda 8: Intensidad = 0,19. Posición = 882,69.



- Banda 9: Intensidad = 0,12. Posición = 919,55.
- Banda 10: Intensidad = 0,18. Posición = 963,69.
- Banda 11: Intensidad = 0,11. Posición = 998,17.
- Banda 12: Intensidad = 0,09. Posición = 1104,19.

El proceso que se sigue para identificar el material es igual que el de la calcita.

En este caso, la lista que se devuelve de los materiales cuyo pico principal se encuentra en el rango 849,69-867,69. es la siguiente:

- Forsterita

En la base de datos el olivino recibe como nombre forsterita por lo que el algoritmo lo ha detectado bien.

Al comparar las bandas de los 2 espectros obtenemos la siguiente información:

nombre	Nº bandas	Bandas	Dif. Intensidad	Dif. Posición
Forsterita	5	5	0,0068	0,0947

*Tabla 5. Resultados de la comparación de bandas*

En los resultados se presencia un problema que va a generar falsos positivos.

En el espectro de la base de datos solo hay 5 bandas del material mientras que en el espectro problema, este material tenía 12 bandas por lo que, al finalizar la primera interacción, después de ordenar la lista y eliminar las bandas utilizadas, nos van a sobrar 7 bandas. El algoritmo va a entender que existe otro material en el espectro y se dispondrá a buscarlo.

Al final de las interacciones del algoritmo, el resultado que se nos presente es de 3 materiales:

- Forsterita (olivino)
- Grosularia
- Jarosita

Se encuentran 2 falsos positivos en este espectro.

Para encontrar el problema hemos hecho uso de la herramienta SpectPro descrita en el apartado 2.2.

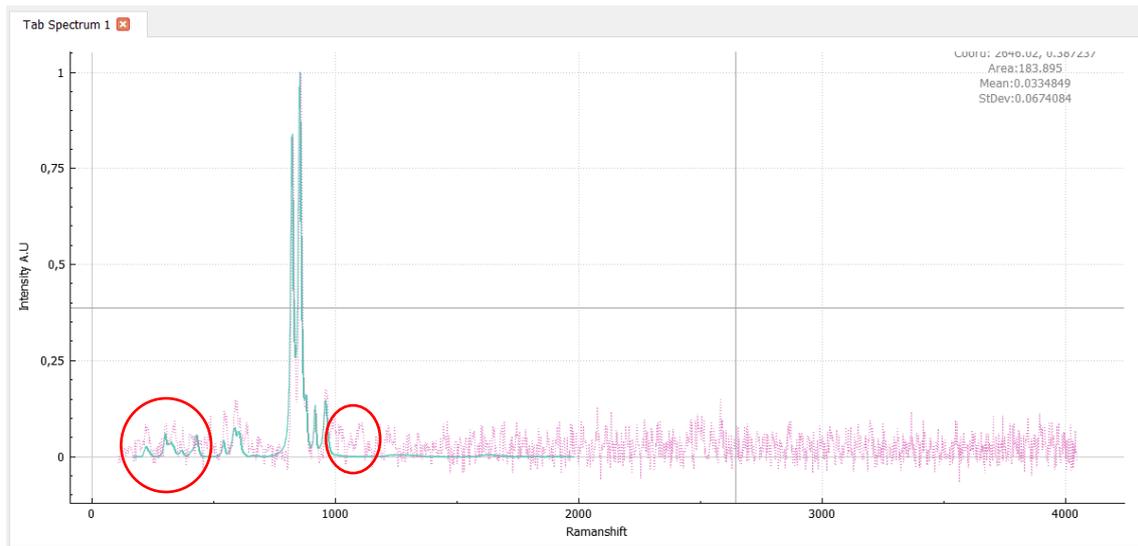


Figura 76. Espectro base de datos y Espectro problema del material olivino

En color verde se representa el espectro del olivino de la base de datos. En el color rosa, el espectro problema.

Las bandas del espectro de la base de datos que hay en el archivo xml que se usa en el algoritmo para solucionar el espectro problema son las siguientes:

- Banda 1: Intensidad = 0,08. Posición = 587.
- Banda 2: Intensidad = 0,84. Posición = 822.
- Banda 3: Intensidad = 1. Posición = 854,50.
- Banda 4: Intensidad = 0,11. Posición = 917,5.
- Banda 5: Intensidad = 0,14. Posición = 960.

En el archivo xml del espectro de la base de datos le faltarían por añadir las bandas que se encuentran señaladas en el lado izquierdo.

Pero el espectro problema también tiene 2 bandas en las posiciones 993 y 1104 que no existen en el espectro original y que a la hora de sacar el archivo xml se han añadido.

El hecho de que estas bandas existan en el espectro problema y no en el espectro de la base de datos ha generado que haya falsos positivos. Por lo tanto, nos hace pensar que, a la hora de probar el código para encontrar la combinación de parámetros óptima, el hecho de que algunos espectros problema vayan a generar falsos positivos independientemente de los parámetros ha influido bastante en los resultados.

### 5.6.3 Cuarzo y Calcita

Las bandas de este espectro problema son:

- Banda 1: Intensidad = 0,33. Posición = 131,45.
- Banda 2: Intensidad = 0,06. Posición = 159,15.



- Banda 3: Intensidad = 0,28. Posición = 208,15.
- Banda 4: Intensidad = 0,06. Posición = 268,15.
- Banda 5: Intensidad = 0,13. Posición = 285,15.
- Banda 6: Intensidad = 0,06. Posición = 359,15.
- Banda 7: Intensidad = 0,04. Posición = 398,15.
- Banda 8: Intensidad = 1. Posición = 468,15.
- Banda 9: Intensidad = 0,04. Posición = 717,15.
- Banda 10: Intensidad = 0,28. Posición = 1090,15.
- Banda 11: Intensidad = 0,03. Posición = 1167,15.

La primera lista de los posibles materiales al buscar materiales con el pico principal en el rango 459,15-477,15 que nos devuelve es la siguiente:

- Sodalita. Su pico principal se encuentra en: 460,5.
- Cuarzo: Su pico principal está en: 465

Al comparar las bandas de estos 2 materiales con las bandas del espectro problema obtenemos las siguientes coincidencias:

nombre	Nº bandas parecidas	Dif. Intensidad	Dif. Posición	Bandas	coincidencia
Sodalita	2	0,012	0,1644	0,2	62,35%
Cuarzo	3	0,0187	0,0626	0	91,87%

*Tabla 6. Resultados de la comparación de las bandas*

El resultado final nos saca como mineral primario el cuarzo que podemos ver que tiene una coincidencia del 91,87%. Podemos ver que aun que la sodalita tenga mayor parecido al espectro problema en las intensidades de las bandas, como el peso de las bandas es mayor que el peso de las intensidades, el cuarzo sale en primer puesto.

Una vez encontrado el material primario, el algoritmo se pone en busca del segundo mineral.

Las bandas que no se han utilizado son:

- Banda 2: Intensidad = 0,2143. Posición = 159,15.
- Banda 4: Intensidad = 0,2143. Posición = 268,15.
- Banda 5: Intensidad = 0,4643. Posición = 285,15.
- Banda 6: Intensidad = 0,2143. Posición = 359,15.
- Banda 7: Intensidad = 0,1429. Posición = 398,15.
- Banda 9: Intensidad = 0,1429. Posición = 717,15.
- Banda 10: Intensidad = 1. Posición = 1090,15.
- Banda 11: Intensidad = 0,1071. Posición = 1167,15.

Las intensidades de las bandas han cambiado debido a la normalización que sufren éstas cada vez que se buscan los minerales.



La lista de los posibles materiales que tengan su pico principal en el rango 1081,15-1099,15 es la siguiente:

- Siderita
- Rodocrosita
- Magnesita
- Epidota
- Calcita
- Aragonita

Después de la comparación se obtienen los siguientes resultados:

nombre	Nº bandas parecidas	Dif. Intensidad	Dif. Posición	Bandas	Coincidencia
Siderita	2	0,0044	0,12	0,3	57,557%
Rodocrosita	3	0,0145	0,0693	0,15	76,6265%
Magnesita	1	0	0,0856	0,45	46,44%
Epidota	2	0,0296	0,1333	0,3	53,709%
Calcita	4	0,0019	0,0811	0	91,69%
Aragonita	1	0,75	0,1517	0,45	29,83%

*Tabla 7. Resultados de la segunda interacción*

En la tabla 7 hay algunos minerales que superan a la calcita en algunos campos como, por ejemplo: Las intensidades de las bandas de la magnesita son los que más se parecen a las del espectro problema; la rodocrosita es la más parecida en cuando a las posiciones de los picos.

Pero la calcita tiene un número de bandas parecidas mayor a los demás.

El resultado nos muestra la calcita en primer lugar por lo que es el elemento secundario del espectro.

Para averiguar el segundo material (calcita) solo se han usado 4 bandas por lo que el algoritmo hará una tercera interacción.

Las bandas que sobran son:

- Banda 4: Intensidad = 1. Posición = 268,15.
- Banda 6: Intensidad = 1. Posición = 359,15.
- Banda 7: Intensidad = 0,66. Posición = 398,15.
- Banda 11: Intensidad = 0,5. Posición = 1167,15.

Como la banda 4 y la banda 6 tenían la intensidad más alta, al volver a normalizar, las 2 bandas pasan a valer 1.

Esto es un posible fallo al generar las bandas del espectro problema. Es muy probable que se haya fallado eligiendo estas bandas confundiéndolas con ruido o que estas



bandas este correctamente seleccionadas, pero no se contemplan en el espectro de la base de datos.

Para conocer el porqué del problema tenemos que observar los espectros de la mezcla y los individuales con la herramienta SpectPro.

Los espectros de los materiales son los siguientes:

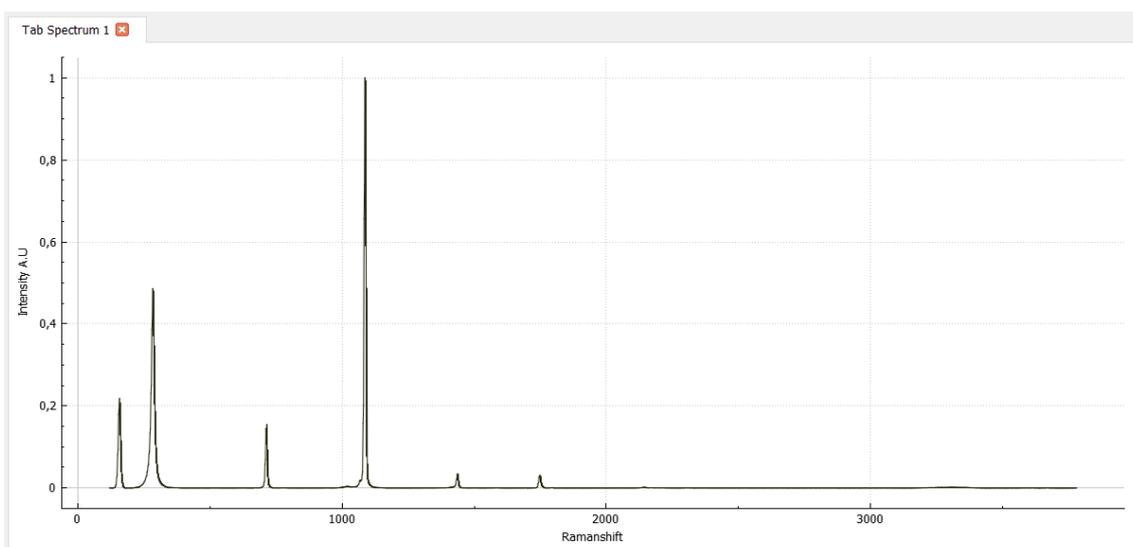


Figura 77. Espectro de la base de datos de la calcita

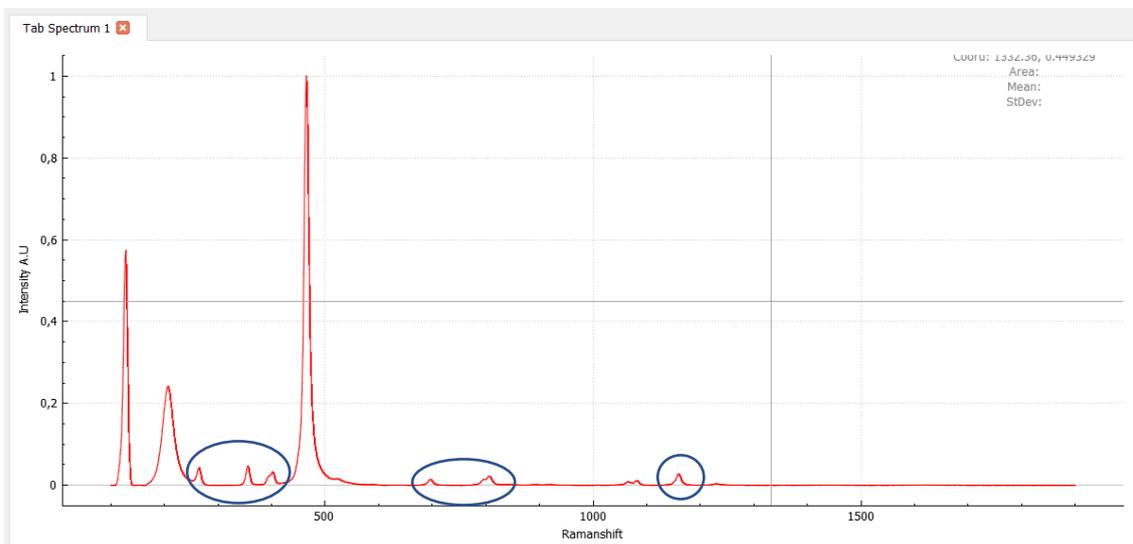


Figura 78. Espectro de la base de datos del cuarzo

En el archivo xml del espectro del cuarzo de la base de datos está compuesto únicamente por 3 bandas como se puede ver en la figura 79, estas bandas son los picos con más intensidad. Como se ve en el espectro, el archivo xml debería tener 9 bandas.

Las bandas que faltan del espectro de la base de datos son las bandas que nos sobran del espectro problema (figura 78) al final de las interacciones del algoritmo.



Gracias a la comprobación de los espectros con la herramienta SpectPro se ha podido ver que este fallo se encuentra en la base de datos con la que se estaba trabajando.

```
<?xml version="1.0" encoding="UTF-8"?>
<ListFittingBands>
  <Band>
    <bandName>Band 1</bandName>
    <bandPosition>127.50</bandPosition>
    <bandIntensity>0.57</bandIntensity>
    <bandFWHM>9.05</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 2</bandName>
    <bandPosition>206.50</bandPosition>
    <bandIntensity>0.24</bandIntensity>
    <bandFWHM>23.92</bandFWHM>
    <lorentzianFlavour>10</lorentzianFlavour>
    <gaussianFlavour>90</gaussianFlavour>
  </Band>
  <Band>
    <bandName>Band 3</bandName>
    <bandPosition>465.00</bandPosition>
    <bandIntensity>1.00</bandIntensity>
    <bandFWHM>11.05</bandFWHM>
    <lorentzianFlavour>15</lorentzianFlavour>
    <gaussianFlavour>85</gaussianFlavour>
  </Band>
</ListFittingBands>
```

Figura 79. Archivo xml del cuarzo

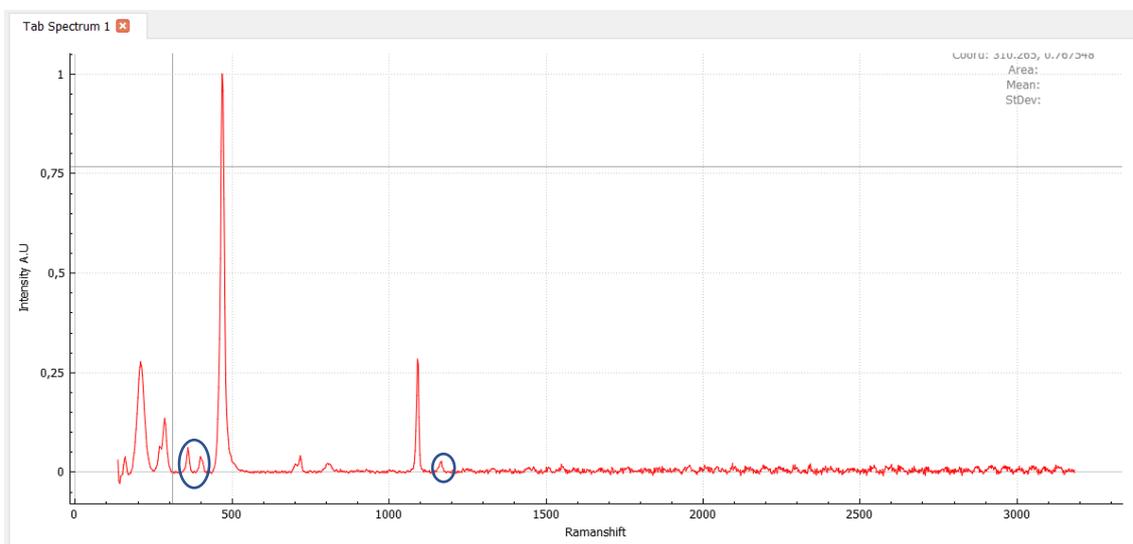


Figura 80. Espectro de la mezcla calcita y cuarzo

Aún con este problema, el algoritmo ha podido dar solución al espectro problema. Si no hubiera 2 bandas con la misma intensidad se habrían generado falsos positivos que habrían afectado a los resultados para la búsqueda de las mejores combinaciones.



## 6 Bibliografía

- [1] "EXOMARS TRACE GAS ORBITER AND SCHIAPARELLI MISSION (2016)," [Online]. Available: <https://exploration.esa.int/web/mars/-/46124-mission-overview>.
- [2] "EXOMARS MISSION (2022)," [Online]. Available: <https://exploration.esa.int/web/mars/-/48088-mission-overview>.
- [3] "CHOOSING THE EXOMARS 2020 LANDING SITE." <https://exploration.esa.int/web/mars/-/53845-landing-site>.
- [4] "THE EXOMARS ROVER INSTRUMENT SUITE." <https://exploration.esa.int/web/mars/-/45103-rover-instruments>.
- [5] F. Rull, "The Raman Laser Spectrometer (RLS) for 2020 Exomars (ESA) Mission: Instrument development and operation on Mars," *J. Medina*, vol. 12, no. 1, 2020.
- [6] F. Rull *et al.*, "The Raman Laser Spectrometer for the ExoMars Rover Mission to Mars," *Astrobiology*, vol. 17, no. 6–7, pp. 627–654, 2017, doi: 10.1089/ast.2016.1567.
- [7] "RLS – The Instrument." <https://auditore.cab.inta-csic.es/rls/rls-the-instrument/>.
- [8] "RLS - RAMAN LASER SPECTROMETER," [Online]. Available: <https://exploration.esa.int/web/mars/-/45103-rover-instruments?section=rls---raman-laser-spectrometer>.
- [9] J. S. Cano *et al.*, "INSTRUMENT DATA ANALYSIS TOOL ( IDAT ) FOR THE ANALYSIS OF RLS DATA," vol. 1, no. 1, pp. 5–6, 2019.
- [10] I. Hermosilla Lopez-Reyes, G., Catala, A., Sanz, A., Llanos, D. R., Rull, F., "Raman spectra processing algorithms and database for RLS-ExoMars," *Eur. Planet. Sci. Congr. EPSC*, vol. 7, pp. 1–2, 2012.