



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

FACULTAD DE INGENIERÍA

**Espectroscopía Raman:  
TRATAMIENTO DE RUIDO ASOCIADO CON  
FLUORESCENCIA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

**Ingeniero Geofísico**

PRESENTA:

**Víctor Manuel Soto Alexander**

DIRECTOR:

Dr. Josué Tago Pacheco



Ciudad Universitaria, Cd. Mx., 2020



*Aurora,*

*Todo el cariño y el amor que me entregaste antes de que la vida nos separara  
ha sido abrigo y calidez durante las noches más frías,  
ha sido escudo contra los momentos más difíciles,  
ha sido la almohada que me transporta a los más hermosos sueños,  
y seguirá cultivando en mí, todos los días,  
a esa persona de la que sé que te sentirías orgullosa.*

*Con todo el amor del mundo,  
Víctor*



# Agradecimientos

---

Primero que nada quiero agradecer a la UNAM y a la Facultad de Ingeniería ya que no importa qué me depare el día de mañana, siempre voy a poder decir con orgullo y con el pecho en alto que pertencí, desde el bachillerato, a una de las mejores universidades del mundo, la máxima casa de estudios de México. La Universidad me ha entregado no solamente una formación académica excelente, me ha dado la oportunidad de conocer a personas maravillosas, amores, relaciones y amigos que espero mantener toda la vida.

A las excelentes personas y profesionales que le dedicaron su tiempo y esfuerzo a la revisión de mi tesis, muchas gracias; M.C. David Escobedo Zenil, M.C. Gerardo Cifuentes Nava, Dr. Aldo Ramos Rosique, M.C. Julián Zapotitla Román, Dr. Giovanni Sosa Ceballos.

Quiero agradecer enormemente al Dr. Josué Tago Pacheco, yo sé que no me van a alcanzar las palabras para decirle gracias por todo el apoyo que me ha brindado. Doy muchas gracias porque tuve la dicha y fortuna de conocerlo como profesor, asesor y compañero. Toda la experiencia, conocimiento y enseñanzas que me llevo, me van a acompañar por el resto de mi vida.

Sería muy ambicioso intentar mencionar cómo estas maravillosas personas han marcado mi camino por la universidad y mi vida, ustedes saben que fueron muy importantes y me gustaría no perderlos nunca; Javier, Nava, Argueta, Roy, Díaz, Karen, Abraham, Yarenni, Aarón, María.

No podría haber llegado aquí sin el apoyo y amor incondicional de mi mamá y de mi papá, espero que me duren toda la vida y hacerlos sentir orgullosos siempre. A mi madrina Coral y a mi tía Marisela, quienes fueron pilares indispensables en mi desarrollo como persona y a quienes siempre les va a pertenecer un lugar especial en mi corazón, les agradezco por todo.

Agradezco la beca recibida a través del proyecto PAPIIT IT102019 en apoyo a mi titulación



# Resumen

---

La Espectroscopía Raman (ER) es una técnica fotónica que permite conocer la estructura cristalina y la composición de los minerales. Una de las complicaciones más comunes al realizar un análisis por ER, es la alteración en el espectro que provoca la fotoluminiscencia de la muestra. Con el fin de identificar con mayor certeza la estructura y la composición de las muestras analizadas, se propone una metodología para disminuir el efecto ruidoso que ocasiona la fotoluminiscencia. A pesar de que la fotoluminiscencia altera la señal hasta volverla irreconocible a simple vista, no destruye el espectro ideal de Raman, solo lo enmascara. Partiendo de esa premisa, se realizaron experimentos bajo diferentes condiciones en el Instituto de Geofísica, Unidad Morelia (IGUM) utilizando un láser monocromático verde de 532 nm.

Para la corrección de la fotoluminiscencia se exploraron los métodos de regresión polinomial, filtro Savitzky-Golay y mínimos cuadrados penalizados con ponderación iterativamente adaptable (airPLS). Con estos métodos, se desarrollaron tres prototipos computacionales en Python que preprocesan los espectros para una mejor identificación de la composición. Se procesaron las señales obtenidas en el IGUM, así como muestras relacionadas con un proyecto del Instituto Nacional de Antropología e Historia (INAH), y se correlacionaron con la base de datos de la Universidad de Arizona (RRUFF). Esta base de datos contiene un catálogo de espectros de Raman de distintos minerales y compuestos ideales. Los resultados demuestran que el preprocesamiento es efectivo y necesario si se pretende encontrar una coincidencia con las bases de minerales ideales. Cabe mencionar que las herramientas de visualización y numéricas programadas en esta tesis, han sido fundamentales para el desarrollo simultáneo del prototipo computacional del proyecto PAPIIT IT102019. Las herramientas desarrolladas se encuentran ahora en uso por el personal del laboratorio del IGUM.





# Índice general

---

Índice de figuras	IX
Índice de tablas	XIII
<b>1. Introducción</b>	<b>1</b>
<b>2. Espectroscopía Raman</b>	<b>3</b>
2.1. Historia . . . . .	3
2.2. Fundamentos teóricos . . . . .	4
2.2.1. Polimorfismo . . . . .	6
2.2.2. Fluorescencia . . . . .	7
2.3. Instrumentación . . . . .	8
<b>3. Metodología para la corrección de espectros</b>	<b>11</b>
3.1. Bases de datos . . . . .	12
3.1.1. Base de datos de referencia (RRUFF) . . . . .	13
3.2. Ajuste de dominios espectrales . . . . .	13
3.3. Corrección de línea base . . . . .	14
3.3.1. Regresión polinomial . . . . .	14
3.3.2. Filtro Savitzky-Golay y señal envolvente . . . . .	18
3.3.3. Mínimos cuadrados penalizados con ponderación iterativa- mente adaptable (airPLS) . . . . .	23
3.4. Filtrado en altas frecuencias . . . . .	27
3.5. Efecto de amplitudes negativas . . . . .	28
3.6. Factor de realce . . . . .	29
3.7. Normalización espectral . . . . .	30
3.8. Correlación espectral . . . . .	30
3.9. Pseudocódigos . . . . .	31

<b>4. Validación de la metodología</b>	<b>35</b>
4.1. Espectros de referencia . . . . .	35
4.2. Regresión polinomial . . . . .	36
4.3. Filtro Savitzky-Golay y señal envolvente . . . . .	41
4.4. Mínimos Cuadrados Penalizados con Ponderación Iterativamente Adaptable (airPLS) . . . . .	44
4.5. Comparación entre las metodologías . . . . .	47
<b>5. Aplicación de la metodología: Proyecto del Instituto Nacional de Antropología e Historia</b>	<b>49</b>
5.1. Los purépecha . . . . .	50
5.2. Muestras arqueológicas . . . . .	51
5.3. Análisis espectral . . . . .	51
5.3.1. Regresión polinomial . . . . .	51
5.3.2. Filtro Savitzky-Golay y señal envolvente . . . . .	55
5.3.3. Mínimos cuadrados penalizados con ponderación iterativa- mente adaptable (airPLS) . . . . .	58
5.4. Comparación entre las metodologías . . . . .	60
<b>6. Conclusiones y comentarios</b>	<b>63</b>
<b>A. Repositorio en GitHub</b>	<b>65</b>
<b>Referencias</b>	<b>67</b>

# Índice de figuras

---

2.1. Espectro electromagnético ordenado de acuerdo al número de onda ( $\tilde{\nu}$ ), la longitud de onda ( $\lambda$ ) y la frecuencia ( $\nu$ ). Tomado de Ferraro (2003). . . . .	5
2.2. Arreglo cristalográfico de dos minerales polimorfos. Izquierda: estructura cristalina del diamante. Derecha: estructura cristalina del grafito. Tomado de Tarbuck et al. (2005) . . . . .	7
2.3. Efecto visual de la fluorescencia en una muestra mineral de lapizlázuli. Izquierda: láser de 254nm. Derecha: láser de 368nm. Tomada de Fosbury (2013). . . . .	8
2.4. Espectro de hidroxipropilmetilcelulosa (HPMC) obtenido con tres láseres de diferente longitud de onda: arriba (514 nm): en medio (633 nm): abajo (785 nm). Tomado de Cebeci-Maltaş (2017) . . . . .	9
2.5. Equipo facilitado por el Laboratorio de Microanálisis del CeMIEGeo	9
3.1. Flujo de la metodología propuesta. . . . .	12
3.2. Factores de ruido que contribuyen a la composición del espectro recolectado. Tomado de Bocklitz et al. (2011). . . . .	14
3.3. Las líneas rojas verticales indican el ancho de banda utilizado en la Figura 3.4, que corresponde al de la base de datos RRUFF. . . . .	17
3.4. Corrección de línea base por regresión polinomial al espectro de turquesa truncado a un ancho de banda cercano a los 1200 [ $cm^{-1}$ ]. El espectro de entrada en todo el ancho de banda se muestra en la Figura 3.3 . . . . .	17
3.5. Cálculo de envolventes a partir de quadratic splines: (a) utilizando una vez el algoritmo; (b) utilizando el algoritmo sobre la envolvente calculada en (a). . . . .	19
3.6. Diferentes iteraciones del filtro Savitzky y Golay aplicado a una señal sintética. . . . .	21
3.7. Filtro Savitzky Golay aplicado a la envolvente del espectro de turquesa. . . . .	22

## ÍNDICE DE FIGURAS

---

3.8. Corrección de línea base por el método de la envolvente y el filtro Savitzky-Golay. . . . .	23
3.9. Diagrama de flujo describiendo el esquema de trabajo del pseudocódigo airPLS. . . . .	26
3.10. Corrección de línea base a un espectro de turquesa por el método de mínimos cuadrados penalizados con ponderación iterativamente adaptable (airPLS). . . . .	27
3.11. Respuesta en frecuencia del filtro Butterworth pasa-bajas. En azul la amplitud de la señal y en verde la frecuencia de corte del filtro Butterworth. . . . .	28
3.12. Ejemplo del efecto del factor de realce sobre una señal. . . . .	29
3.13. Resultado de aplicar la metodología propuesta con airPLS a un espectro de Hematita. . . . .	31
4.1. Resultado de aplicar la metodología de regresión polinomial al espectro de una muestra de turquesa. . . . .	37
4.2. Correlación entre los espectros de referencia de una aheylita y de una turquesa. . . . .	38
4.3. Resultado de aplicar la metodología de regresión polinomial al espectro de una muestra de jarosita. . . . .	39
4.4. Resultado de aplicar la metodología de regresión polinomial al espectro de una muestra de dickita. . . . .	40
4.5. Resultado de aplicar la metodología de la envolvente y el filtro Savitzky-Golay al espectro de una muestra de turquesa. . . . .	41
4.6. Resultado de aplicar la metodología de la envolvente y el filtro Savitzky-Golay al espectro de una muestra de jarosita. . . . .	43
4.7. Resultado de aplicar la metodología de la envolvente y el filtro Savitzky-Golay al espectro de una muestra de dickita. . . . .	44
4.8. Resultado de aplicar la metodología airPLS al espectro de una muestra de turquesa. . . . .	45
4.9. Resultado de aplicar la metodología airPLS al espectro de una muestra de jarosita. . . . .	46
4.10. Resultado de aplicar la metodología airPLS al espectro de una muestra de dickita. . . . .	47
4.11. Representación del tiempo de cómputo necesario y el coeficiente de correlación obtenido para las muestras analizadas mediante cada método. . . . .	48
5.2. Espectro corregido de bezote de obsidiana con incrustación . . . .	51
5.1. Símbolos de poder purépecha . . . . .	52

5.3. Corrección del espectro de los pendientes de jadeíta por el método de regresión polinomial. . . . .	53
5.4. Corrección del espectro del sartal de turquesas por el método de regresión polinomial. . . . .	54
5.5. Corrección del espectro del bezote de obsidiana con incrustación por el método de filtro Savitzky-Golay. . . . .	55
5.6. Corrección del espectro de los pendientes de jadeíta por el método de filtro Savitzky-Golay. . . . .	56
5.7. Corrección del espectro del sartal de turquesas por el método de filtro Savitzky-Golay. . . . .	57
5.8. Corrección del espectro del bezote de obsidiana con incrustación por el método airPLS. . . . .	58
5.9. Corrección del espectro de los pendientes de jadeíta por el método airPLS. . . . .	59
5.10. Corrección del espectro del sartal de turquesas por el método airPLS. . . . .	60
5.11. Representación del tiempo de cómputo necesario y el coeficiente de correlación obtenido para las muestras analizadas mediante cada método. . . . .	61



# Índice de tablas

---

2.1. Regiones espectrales y su origen. Tomado de Ferraro (2003) . . . .	6
4.1. Coeficientes de correlación más altos como resultado de la metodología de regresión polinomial al espectro de una dickita. . . . .	40
4.2. Coeficientes de correlación más altos como resultado de la metodología de la envolvente y el filtro Savitzky-Golay al espectro de una turquesa. . . . .	42





---

## Capítulo 1

# Introducción

---

Cuando un láser de luz monocromático (i.e. una sola frecuencia) atraviesa un medio (gas, líquido o sólido) la mayor parte de la luz dispersada permanecerá en la frecuencia del láser. Sin embargo, una pequeña porción de la luz dispersada cambiará de frecuencia, por encima y por debajo de la frecuencia incidente, a esto es a lo que se le conoce como efecto Raman. El efecto Raman fue observado por primera vez por [Raman y Krishnan \(1928\)](#) utilizando luz solar concentrada y filtros, basándose en las observaciones de los cambios de color en la luz dispersada. No fue sino hasta la llegada de los láseres visibles de onda continua, durante la década de 1960, que la importancia de la Espectroscopía Raman fue considerada como una técnica analítica de rutina. Además, la capacidad de esta fuente de luz monocromática altamente intensa, que podía concentrarse en un área reducida, permitió el análisis de pequeños volúmenes de gas, líquido o sólido ([Potts et al., 1995](#)).

Instrumentos de Raman son producidos hoy en día por diferentes fabricantes, lo que permite la aplicación de la espectroscopía Raman a pequeñas muestras como fibras o partículas de polvo. Dentro de las Ciencias de la Tierra, la Espectroscopía Raman representa una técnica analítica relativamente nueva, que a la fecha solo se utiliza en unos cuantos laboratorios geológicos. Los aspectos particularmente atractivos de la técnica Raman son la relativa falta de preparación de la muestra y la naturaleza no destructiva del análisis. Estos atributos han resultado en una variedad de aplicaciones geológicas, un ejemplo es la caracterización y el estudio estructural de los minerales ([McMillan, 1989](#)).

Lamentablemente, en muchas ocasiones, el espectro es ruidoso, lo que complica el análisis y dificulta la labor del analista químico para encontrar características de interés. En muchas formas de espectroscopía [e.g. resonancia magnética nuclear (NMR, por sus siglas en inglés)] se utiliza la primera o segunda derivada del espectro, pero estos espectros solamente pueden utilizarse en espectros que tengan una alta relación señal-ruido (SNR, por sus siglas en inglés) debido a que

## 1. INTRODUCCIÓN

---

los espectros derivados contienen una proporción más grande de ruido en altas frecuencias (Konorov et al., 2006). Además, si el espectro va a ser interpretado por una computadora, por ejemplo para buscar el espectro en una base de datos de referencia, el ruido del espectro sin procesamiento, en muchas ocasiones puede llevar a pobres resultados de búsqueda. Ciertamente, remover el ruido de un espectro es de gran interés práctico (Rowlands y Elliott, 2011).

Por iniciativa del Dr. Giovanni Sosa Ceballos del Instituto de Geofísica, Unidad Morelia (IGUM), y a través del Dr. Josué Tago Pacheco, recibí la invitación para atacar el problema que presentan colaboradores del Laboratorio de Microanálisis del Centro Mexicano de Innovación en Energía Geotérmica (CEMIEGeo).

El objetivo último de la tesis es el de desarrollar un programa automatizado que elimine lo mejor posible el ruido de los espectros de Raman obtenidos en el IGUM, y que compare el espectro corregido con una base de datos de espectros de referencia. Este procesamiento mejora significativamente la clasificación que se obtendría cuando se utilizan los espectros sin corrección.

La principal contribución de la tesis es que el código generado queda a disposición del IGUM y de cualquier dependencia de la UNAM que pudiera llegar a presentar un problema de naturaleza similar.

Esta tesis se divide en 6 capítulos: Al principio se encuentra la introducción. El capítulo 2 contiene el marco teórico del Efecto Raman, así como las especificaciones del equipo utilizado para recolectar los espectros analizados. En el capítulo 3 se presenta la metodología propuesta para el tratamiento de los espectros del IGUM. El capítulo 4 ilustra, con un espectro real como ejemplo, el impacto de la metodología propuesta. El capítulo 5 muestra la aplicación del prototipo computacional desarrollado en muestras recolectadas por el Instituto Nacional de Antropología e Historia (INAH), con la colaboración del Museo Regional Michoacán. Finalmente, el capítulo 6 contiene las conclusiones del presente trabajo.

---

## Capítulo 2

# Espectroscopía Raman

---

<b>2.1. Historia</b> . . . . .	<b>3</b>
<b>2.2. Fundamentos teóricos</b> . . . . .	<b>4</b>
2.2.1. Polimorfismo . . . . .	6
2.2.2. Fluorescencia . . . . .	7
<b>2.3. Instrumentación</b> . . . . .	<b>8</b>

---

En este capítulo se presentan los fundamentos teóricos sobre los que se sustenta el método de espectroscopía Raman como técnica óptica para la clasificación de minerales. Además, se aborda la importancia de diferenciar la estructura cristalina de las especies moleculares polimórficas. Por otro lado, se incluye una breve descripción del equipo facilitado por el Laboratorio de Microanálisis del Centro Mexicano de Innovación en Energía Geotérmica (CeMIEGeo). Por último, se introduce el efecto de la fotoluminiscencia y sus efectos en el espectro de Raman.

## 2.1. Historia

El fenómeno de dispersión inelástica de la luz fue primero postulado por **Smekal (1923)** y observado experimentalmente en 1928 por **Raman y Krishnan**. Desde entonces se refiere al fenómeno como espectroscopía Raman. En el experimento original la luz solar se concentraba mediante un telescopio en una muestra que podía ser líquido purificado o vapor libre de polvo. Un segundo lente se colocaba cercano a la muestra para coleccionar la radiación dispersada. Un sistema de filtros ópticos se usaba para mostrar la existencia de radiación dispersada con

una frecuencia alterada de la luz incidente, lo que es la característica básica de la espectroscopía Raman.

### 2.2. Fundamentos teóricos

Cuando la luz interactúa con la materia, los fotones que componen la luz puede que sean absorbidos o dispersados, o pueden no interactuar con el material y pasar a través de él. Si la energía de un fotón incidente coincide con la diferencia de energía entre el estado basal de una molécula y su estado excitado, el fotón puede ser absorbido y la molécula ascender a un estado excitado de alta energía. Es este cambio de estado que es medido en la espectroscopía de absorción, mediante la detección de la pérdida de esa energía de radiación de la luz. Sin embargo, también es posible que el fotón interactúe con la molécula y se disperse de ella. En este caso no hay necesidad de que el fotón tenga una energía igual a la diferencia entre dos niveles de energía de la molécula. Los fotones dispersados pueden ser observados recogiendo la luz en ángulo con respecto al haz de luz incidente, y siempre que no haya absorción de ninguna transmisión electrónica que tenga energías similares a la de la luz incidente.

La dispersión es una técnica usada comúnmente. Por ejemplo, es ampliamente usada para medir el tamaño de partículas para tamaños por debajo de  $1 \mu m$ .

Un ejemplo de todos los días, es que el cielo es azul debido a que la luz azul de alta energía es dispersada de las partículas y moléculas en la atmósfera más eficientemente que la luz roja de baja energía. No obstante, la principal técnica de dispersión usada para la identificación molecular es la dispersión Raman.

El proceso de absorción es usado en un amplio rango de técnicas espectroscópicas. Por ejemplo se usa en espectroscopía acústica donde existe una diferencia de energía pequeña entre el estado basal y excitado, y en espectroscopía de rayos-x donde existe una diferencia muy grande.

La radiación se caracteriza usualmente por su longitud de onda ( $\lambda$ ). Sin embargo, en espectroscopía, como nos interesa la interacción de la radiación con estados de las moléculas y estos usualmente se discuten en términos de energía, es conveniente utilizar escalas de frecuencia ( $\nu$ ) o número de onda ( $\bar{\omega}$ ), que se relacionan linealmente con la energía. Las relaciones entre estas escalas se dan a continuación:

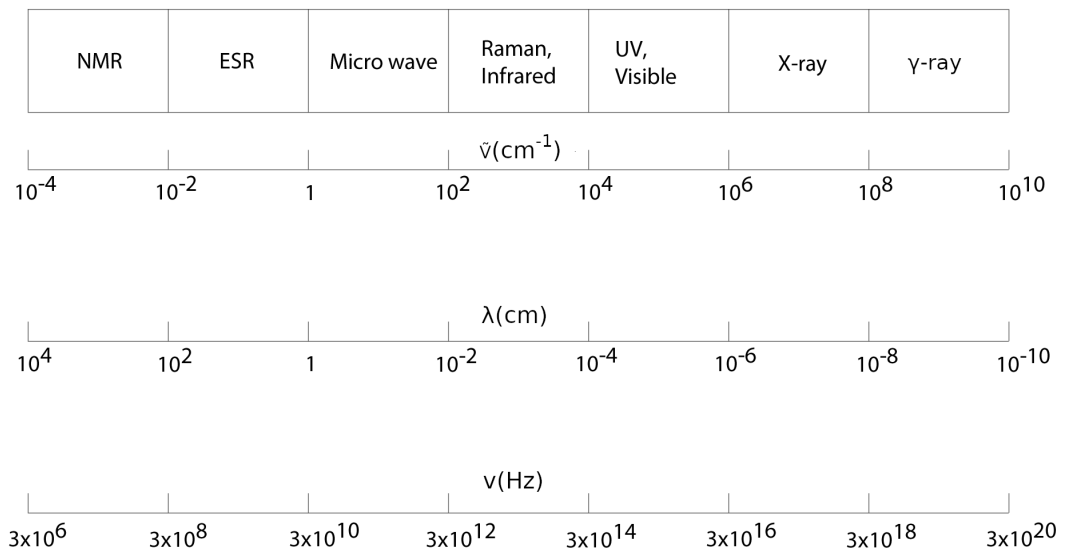
$$\lambda = c/\nu \tag{2.1}$$

$$\nu = \Delta E/h \tag{2.2}$$

$$\bar{\omega} = v/c = 1/\lambda \quad (2.3)$$

donde  $c$  es la velocidad de la luz,  $\Delta E$  es la energía del fotón y  $h$  es la constante de Planck.

Es claro de las ecuaciones 2.1-2.3 que la energía es inversamente proporcional a la longitud de onda y por lo tanto, la región de alta energía se encuentra a la derecha de la figura 2.1 y las longitudes de onda más grandes se encuentran a la izquierda.



**Figura 2.1:** Espectro electromagnético ordenado de acuerdo al número de onda ( $\tilde{\nu}$ ), la longitud de onda ( $\lambda$ ) y la frecuencia ( $\nu$ ). Tomado de Ferraro (2003).

La manera en la que la radiación es empleada en espectroscopía Raman e infrarroja es diferente. En la espectroscopía infrarroja, la energía infrarroja, que cubre un rango de frecuencias, (ver Tabla 2.1 sobre regiones espectrales como función del número de onda) es dirigida a la muestra. La absorción ocurre cuando la frecuencia de la radiación incidente concuerda con la vibración que eleva a la molécula a un estado excitado. En la espectroscopía Raman se utiliza una sola frecuencia de radiación para irradiar la muestra y es la radiación dispersada de la molécula, a una diferente frecuencia de la incidente, la que se detecta.

## 2. ESPECTROSCOPIA RAMAN

---

**Tabla 2.1:** Regiones espectrales y su origen. Tomado de Ferraro (2003)

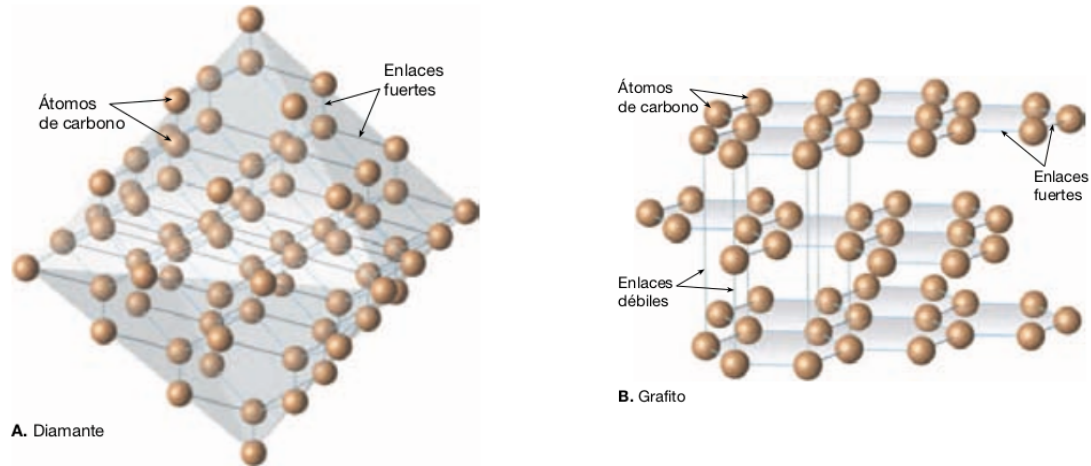
Espectroscopía	Rango ( $\bar{\omega}$ , $cm^{-1}$ )	Origen
Rayos $\gamma$	$10^{10} - 10^8$	Reacomodo de partículas elementales en el núcleo
Rayos $X$	$10^8 - 10^6$	Transición entre niveles de energía de los electrones internos de átomos
UV - Visible	$10^6 - 10^4$	Transición entre niveles de energía de los electrones de valencia de átomos
Raman e Infrarrojo	$10^4 - 10^2$	Transición entre niveles vibracionales (cambio de configuración)
Microondas	$10^2 - 1$	Transición entre niveles rotacionales (cambio de orientación)
Electron spin resonance	$1 - 10^{-2}$	Transición entre niveles de spin en los electrones en el campo magnético
Nuclear magnetic resonance	$10^{-2} - 10^{-4}$	Transición entre niveles de spin de los núcleos en campos magnéticos

### 2.2.1. Polimorfismo

Si bien es cierto que dos muestras de un mismo mineral presentan la misma estructura cristalina, algunos elementos son capaces de formar varias formas cristalinas. Es por esto que dos minerales con propiedades completamente diferentes pueden tener exactamente la misma composición química. Minerales de este tipo se conocen como polimorfos Tarbuck et al. (2005). Un ejemplo de polimorfismo se encuentra en el diamante y en el grafito ya que ambos se constituyen exclusivamente de carbono, y presentan características considerablemente diferentes debido a su arreglo cristalino (ver Figura 2.2). La diferenciación de polimorfos es importante, particularmente en el área de la geología. Diferentes especies en una serie de compuestos polimorfos se pueden diferenciar con espectroscopía Raman. Esto, debido a que los espectros de Raman consisten en bandas o picos que dependen de los modos vibracionales de la red cristalina.

### Soluciones sólidas

Otro ejemplo de cómo ayuda la espectroscopía Raman a detectar cambios estructurales dentro de especies minerales se encuentra analizando las soluciones sólidas. Las plagioclasas (feldespatos) constituyen el ejemplo más común. El contenido de sodio ( $Na$ ) y calcio ( $Ca$ ) varía entre los miembros anortita y albita respectivamente. Aunque toda la solución sólida es triclinica, el cambio en



**Figura 2.2:** Arreglo cristalográfico de dos minerales polimorfos. Izquierda: estructura cristalina del diamante. Derecha: estructura cristalina del grafito. Tomado de Tarbuck et al. (2005)

composición genera una distorsión espectral sutil. En contraste, los feldespatos potásicos son monoclínicos, así que este cambio de estructura ayuda a diferenciar mejor entre ambas especies minerales debido al cambio en el espectro de Raman.

### 2.2.2. Fluorescencia

La fluorescencia inducida por láser es la fuente más común de interferencia en las mediciones de Raman. Por lo tanto, para lograr un uso adecuado de la espectroscopía Raman, es necesario extraer firmas de Raman de esta radiación.

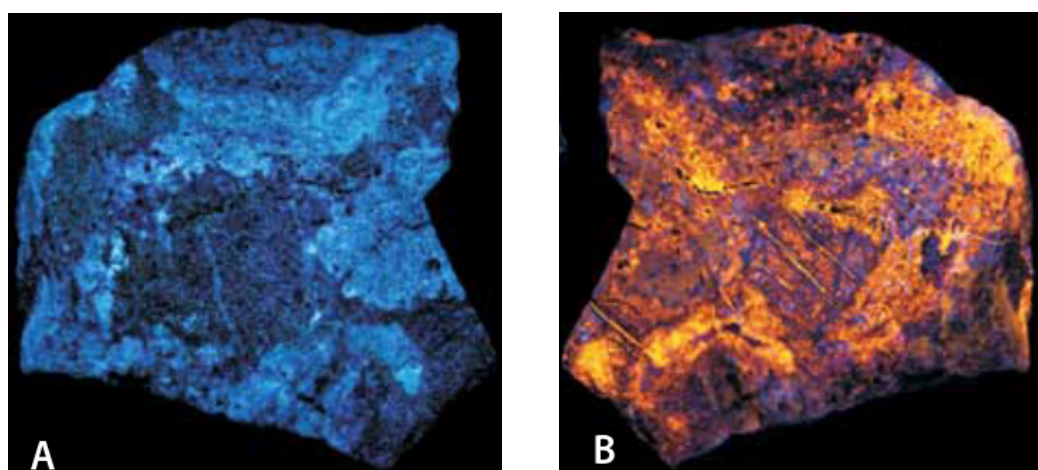
Cuando un fotón de luz con suficiente energía interactúa con un átomo o una molécula, un electrón puede excitarse desde el nivel más bajo, el estado fundamental, a un estado excitado. Este electrón a veces decae casi inmediatamente de regreso al lugar donde comenzó emitiendo un fotón de la misma longitud de onda. Esto se llama fluorescencia de resonancia y es interesante para los físicos cuánticos, pero menos para los mineralogistas. Sin embargo, el electrón puede hacer cosas más exóticas. Dentro de un cristal, puede rebotar un poco, emitiendo un poco de calor en el camino, para alcanzar un estado de excitación más bajo desde el cual puede hacer una transición radiativa a un estado aún más bajo y posiblemente fundamental al emitir un fotón más rojo que el incidente (ver Fig. 2.3). En una molécula puede haber un proceso análogo en el que el electrón excitado puede decaer primero al nivel vibracional más bajo del estado excitado y luego hacer una transición, emitiendo un fotón, a uno de los niveles vibracionales

## 2. ESPECTROSCOPIA RAMAN

---

asociados con el estado fundamental Fosbury (2013).

La fluorescencia puede ocurrir por el compuesto analizado o por impurezas fluorescentes en la muestra. Este efecto de la fluorescencia depende en gran medida de la longitud de onda del láser incidente, por el contrario, el corrimiento de Raman no varía cambiando la frecuencia del láser de excitación. Láseres visibles como el rojo o el verde de 532 y 633 nm respectivamente, generan un efecto de fotoluminiscencia más elevado, debido a que proporcionan fotones más energéticos para producir la excitación de las moléculas de la muestra. Conforme el láser se aproxima a la región cercana al infrarrojo con láseres de 785 nm o 1064 nm el efecto de la fluorescencia se va atenuando o completamente desaparece debido a que la energía de estas longitudes de onda puede no ser suficiente para excitar las moléculas a un estado electrónico superior. Este efecto se puede observar en la Figura 2.4

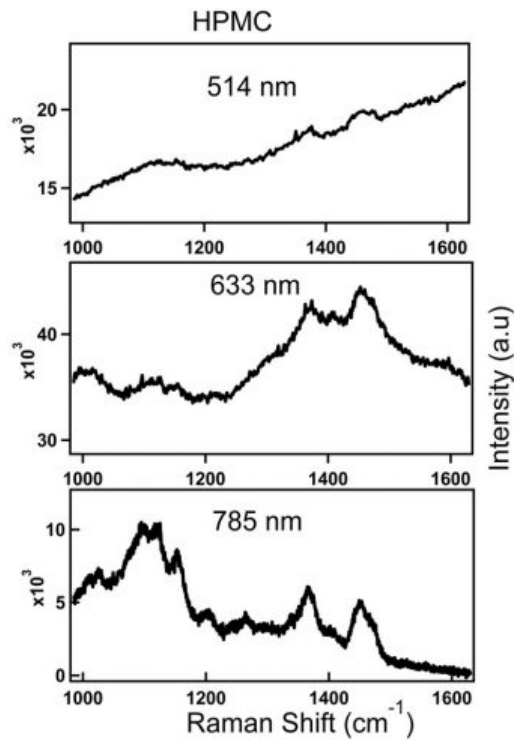


**Figura 2.3:** Efecto visual de la fluorescencia en una muestra mineral de lapizlázuli. Izquierda: láser de 254nm. Derecha: láser de 368nm. Tomada de Fosbury (2013).

### 2.3. Instrumentación

El espectroscopio de Raman que se utilizó fue un microscopio Thermo Scientific™ DXR™ (ver Figura 2.5a). Este equipo realiza la identificación de fases minerales y compuestos orgánicos sin la necesidad pero con la posibilidad de montar la muestra en láminas delgadas. Entre otras actividades, determina concentraciones de H<sub>2</sub>O y CO<sub>2</sub> disueltos en vidrios volcánicos. El espectroscopio Raman cuenta con dos láseres, uno de 532 nm y otro de 780 nm. El láser utilizado para este proyecto fue el láser verde de 532 nm que se muestra en la Figura 2.5b.





**Figura 2.4:** Espectro de hidroxipropilmetilcelulosa (HPMC) obtenido con tres láseres de diferente longitud de onda: arriba (514 nm): en medio (633 nm): abajo (785 nm). Tomado de [Cebeci-Maltaş \(2017\)](#)



(a) Espectroscopio Raman



(b) Láser de 532 nm

**Figura 2.5:** Equipo facilitado por el Laboratorio de Microanálisis del CeMIEGeo



---

## Capítulo 3

# Metodología para la corrección de espectros

---

<b>3.1. Bases de datos</b> . . . . .	<b>12</b>
3.1.1. Base de datos de referencia (RRUFF) . . . . .	13
<b>3.2. Ajuste de dominios espectrales</b> . . . . .	<b>13</b>
<b>3.3. Corrección de línea base</b> . . . . .	<b>14</b>
3.3.1. Regresión polinomial . . . . .	14
3.3.2. Filtro Savitzky-Golay y señal envolvente . . . . .	18
3.3.3. Mínimos cuadrados penalizados con ponderación iterativamente adaptable (airPLS) . . . . .	23
<b>3.4. Filtrado en altas frecuencias</b> . . . . .	<b>27</b>
<b>3.5. Efecto de amplitudes negativas</b> . . . . .	<b>28</b>
<b>3.6. Factor de realce</b> . . . . .	<b>29</b>
<b>3.7. Normalización espectral</b> . . . . .	<b>30</b>
<b>3.8. Correlación espectral</b> . . . . .	<b>30</b>
<b>3.9. Pseudocódigos</b> . . . . .	<b>31</b>

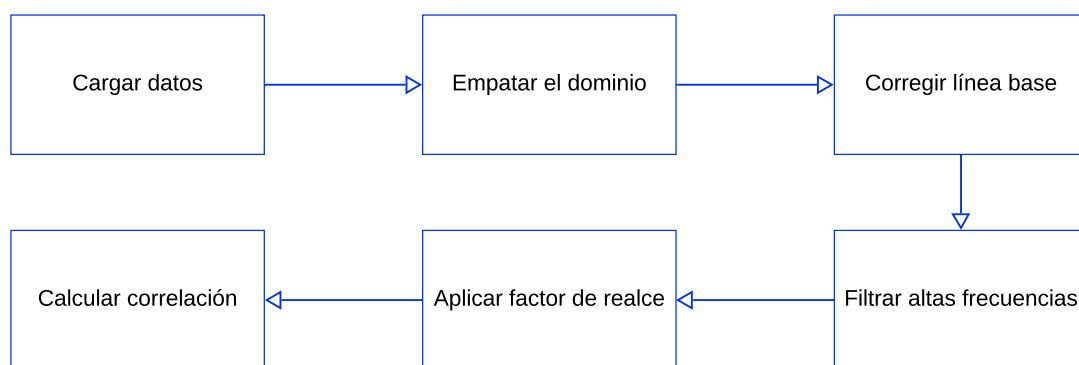
---

### 3. METODOLOGÍA PARA LA CORRECCIÓN DE ESPECTROS

---

En este capítulo se muestra, con detalle, cada uno de los componentes de la metodología propuesta para corregir el efecto de la fotoluminiscencia. Además, se presenta la metodología utilizada para correlacionar los espectros corregidos con los de la base de datos.

El procedimiento general de la metodología se ilustra en el flujo de la Figura 3.1.



**Figura 3.1:** Flujo de la metodología propuesta.

#### 3.1. Bases de datos

En el mundo de hoy, muchas personas confían en la disponibilidad de bases de datos para realizar actividades diarias como verificar el pronóstico meteorológico. Este es un ejemplo de acción que cualquiera puede realizar de forma rápida, rutinaria y, a menudo, a bajo costo, debido al acceso a Internet y al desarrollo de bases de datos extensas.

La comunidad científica siempre se ha basado en bases de datos, que en los primeros años solo estaban disponibles a través de recopilaciones de revistas, libros y registros de datos. La naturaleza analógica de estos recursos de datos limita drásticamente el proceso de búsqueda de registros y el establecimiento de relaciones entre conjuntos de datos. Hoy en día, las bases de datos científicas digitales y la World Wide Web han cambiado la forma en que podemos hacer ciencia. Esta nueva generación de bases de datos no solo proporciona acceso a la información, sino que permite el intercambio de datos entre proyectos de investigación multidisciplinarios y hace posible el enlace cruzado de diversos tipos de datos.

Los minerales son muy importantes tanto para la economía mundial como para la investigación científica, suministran las materias primas necesarias para el desarrollo de las sociedades modernas y su estudio proporciona conocimiento

de nuestro planeta, el sistema solar y la naturaleza misma del universo. Como tal, su identificación y caracterización son críticas para muchos campos. Sin embargo, solo unos pocos expertos pueden identificar con precisión los minerales sin instrumentos analíticos y sus bases de datos complementarias. En general, los minerales se caracterizan de manera única por la química y la estructura cristalina, o sus propiedades derivadas. Por ejemplo, la propiedad derivada, el color, puede ser una función de la química o el hábito cristalino una función de la estructura.

### 3.1.1. Base de datos de referencia (RRUFF)

La identificación de minerales mediante la espectroscopía Raman se realiza normalmente mediante rutinas de búsqueda/coincidencia que comparan el espectro adquirido con los espectros de referencia de una base de datos. El propósito del proyecto RRUFF es desarrollar una base de datos Raman de este tipo midiendo la química, los patrones de difracción de rayos X, Raman y los espectros infrarrojos de los minerales conocidos y hacer que estos datos estén disponibles de forma fácil y gratuita para la comunidad científica, la industria y el público general (Bindi et al., 2015).

Por ser una base de datos muy completa (más de 4000 muestras) y de libre acceso, se eligió para descargar los espectros. Los archivos de la base de datos RRUFF se pueden descargar de [https://rruff.info/zippered\\_data\\_files/raman/excellent\\_unoriented.zip](https://rruff.info/zippered_data_files/raman/excellent_unoriented.zip).

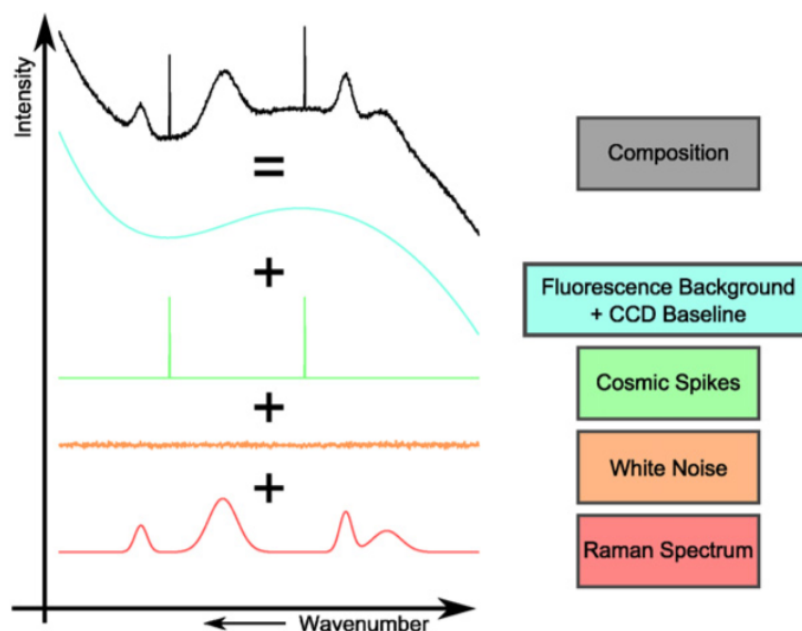
El archivo que se descarga del enlace anterior es un archivo comprimido que contiene archivos de texto (*.txt*) con el nombre de la muestra que se analizó. En la primera columna de cada archivo de texto se encuentra el número de onda y en la segunda columna, la intensidad de la señal de Raman.

## 3.2. Ajuste de dominios espectrales

Cada espectro de la base de datos RRUFF presenta un rango distinto de número de onda (i.e. 1000-2300  $\bar{\omega}$ , 800-4000  $\bar{\omega}$ ). Dado que el espectro a corregir y los espectros de RRUFF no comparten el mismo ancho de banda, (es decir, no empiezan y terminan en el mismo número de onda) es necesario comenzar por atender este problema. Esto, con la finalidad de únicamente corregir el espectro recolectado en el ancho de banda correspondiente con cada espectro de RRUFF. Un ejemplo de este proceso puede observarse en las Figuras 3.3 y 3.4.

### 3.3. Corrección de línea base

Como se ha mencionado anteriormente, los espectros de Raman suelen presentar ruido que complica su correcta clasificación (ver Figura 3.2). Uno de los componentes más fuertes dentro de este ruido es la fluorescencia. Este efecto altera, en algunas ocasiones, fuertemente la línea base del espectro. Existen diversos métodos para intentar disminuir estas desviaciones de la línea base; algunas de ellas las presentaremos a continuación.



**Figura 3.2:** Factores de ruido que contribuyen a la composición del espectro recolectado. Tomado de [Bocklitz et al. \(2011\)](#).

#### 3.3.1. Regresión polinomial

Como una primera aproximación al problema de la línea base, se utiliza un ajuste polinomial por mínimos cuadrados ([Montgomery et al., 2012](#)). Existen diversos métodos para elaborar este ajuste, dos de ellos se comentan a continuación.

Se parte del supuesto de que la desviación de la línea base se puede aproximar como una curva polinomial suave. Se tiene la ecuación general para un polinomio de grado  $k$

$$y = a_0 + a_1x + \dots + a_kx^k. \quad (3.1)$$

Lo que se pretende con el método de mínimos cuadrados es encontrar el polinomio de grado  $k$  que ajuste los datos del espectro con el mínimo error. Se utiliza la función residual

$$R^2 \equiv \sum_{i=1}^n [y_i - (a_0 + a_1x_i + \dots + a_kx_i^k)]^2. \quad (3.2)$$

De la ecuación 3.2 se observa que se eleva al cuadrado el error para evitar que la aportación de errores con valores positivos y negativos se anule. Ya que esto podría arrojar un falso residual.

Para encontrar los coeficientes del polinomio que alcancen el error mínimo, se procede a derivar parcialmente la función del residual con respecto a los coeficientes  $(a_0, a_1, \dots, a_k)$  y se igualan con cero (Ver ecuación 3.3).

$$\begin{aligned} \frac{\partial(R^2)}{\partial a_0} &= -2 \sum_{i=1}^n [y_i - (a_0 + a_1x_i + \dots + a_kx_i^k)] = 0, \\ \frac{\partial(R^2)}{\partial a_1} &= -2 \sum_{i=1}^n [y_i - (a_0 + a_1x_i + \dots + a_kx_i^k)] x_i = 0, \\ &\vdots \\ \frac{\partial(R^2)}{\partial a_k} &= -2 \sum_{i=1}^n [y_i - (a_0 + a_1x_i + \dots + a_kx_i^k)] x_i^k = 0. \end{aligned} \quad (3.3)$$

Esto lleva a las ecuaciones

$$\begin{aligned} a_0n + a_1 \sum_{i=1}^n x_i + \dots + a_k \sum_{i=1}^n x_i^k &= \sum_{i=1}^n y_i, \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 + \dots + a_k \sum_{i=1}^n x_i^{k+1} &= \sum_{i=1}^n x_i y_i, \\ &\vdots \\ a_0 \sum_{i=1}^n x_i^k + a_1 \sum_{i=1}^n x_i^{k+1} + \dots + a_k \sum_{i=1}^n x_i^{2k} &= \sum_{i=1}^n x_i^k y_i, \end{aligned} \quad (3.4)$$

o en forma matricial

$$\begin{bmatrix} n & \sum_{i=1}^n x_i & \cdots & \sum_{i=1}^n x_i^k \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \cdots & \sum_{i=1}^n x_i^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_i^k & \sum_{i=1}^n x_i^{k+1} & \cdots & \sum_{i=1}^n x_i^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \vdots \\ \sum_{i=1}^n x_i^k y_i \end{bmatrix}. \quad (3.5)$$

La única incógnita de la ecuación 3.5 es el vector de coeficientes polinomiales  $(a_0, a_1, \dots, a_k)$ ; por lo tanto, al resolver la ecuación matricial, se obtienen los coeficientes buscados. La matriz de la ecuación 3.5 es una matriz de Vandermonde, (Weisstein, 2002), que también se puede obtener evaluando a los polinomios en

### 3. METODOLOGÍA PARA LA CORRECCIÓN DE ESPECTROS

---

los datos como

$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \quad (3.6)$$

Al premultiplicar ambos lados por la transpuesta de la primera matriz se obtiene

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \quad (3.7)$$

Desarrollando

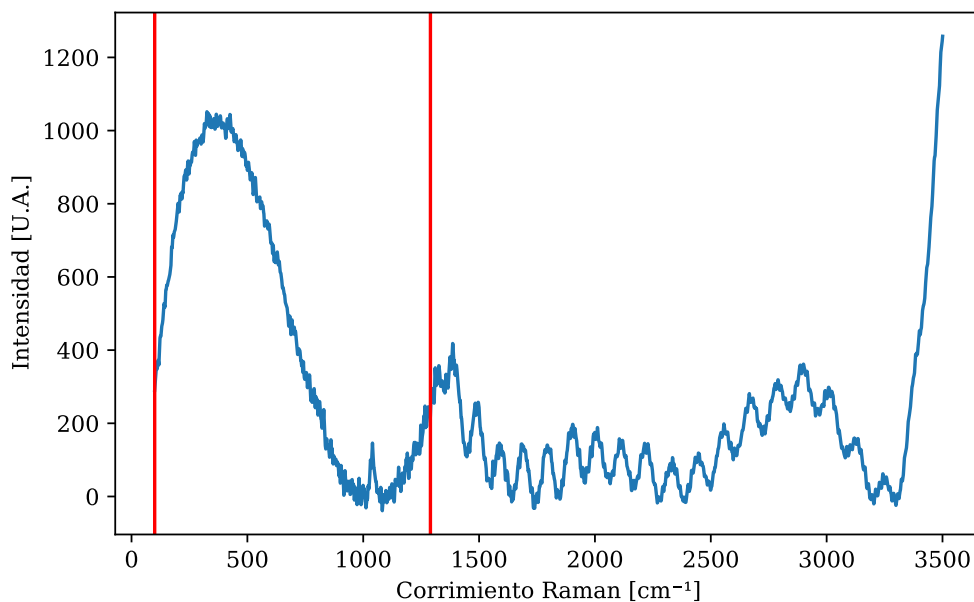
$$\begin{bmatrix} n & \sum_{i=1}^n x_i & \cdots & \sum_{i=1}^n x_i^k \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \cdots & \sum_{i=1}^n x_i^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_i^k & \sum_{i=1}^n x_i^{k+1} & \cdots & \sum_{i=1}^n x_i^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \vdots \\ \sum_{i=1}^n x_i^k y_i \end{bmatrix}. \quad (3.8)$$

En forma matricial

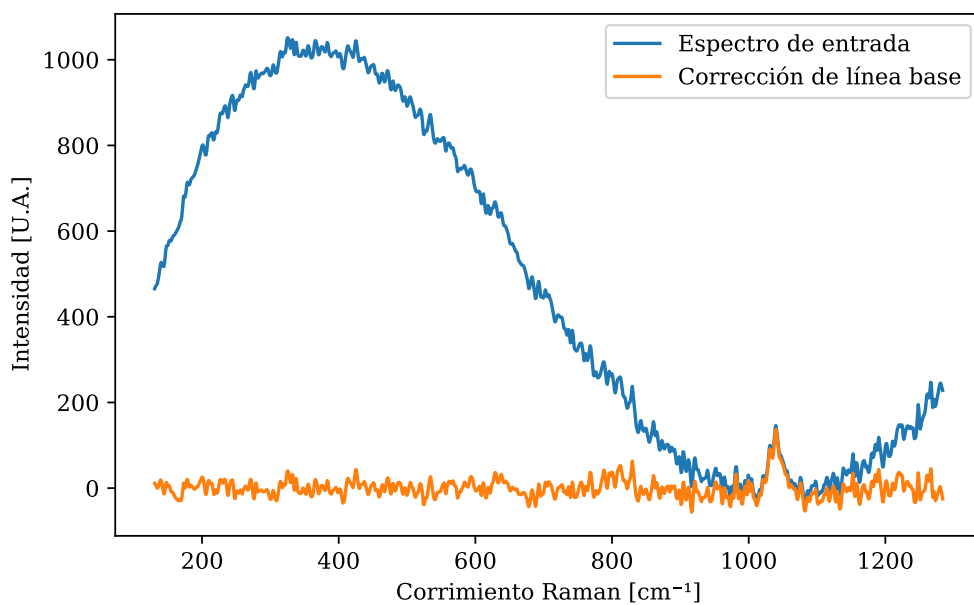
$$\begin{aligned} Xa &= y \\ X^T Xa &= X^T y \\ a &= (X^T X)^{-1} X^T y \end{aligned}. \quad (3.9)$$

Para ilustrar el resultado de la regresión polinomial, se utiliza el espectro de turquesa mostrado en la Figura 3.3. En este espectro se puede observar el fuerte efecto del ruido asociado con fluorescencia. En la Figura 3.4 se muestra la corrección de línea base por regresión polinomial, así como el espectro de la turquesa después de corregir los dominios. A pesar de que la señal corregida disminuyó significativamente el efecto de la fluorescencia, aún persiste el ruido en altas frecuencias, problema que se discute más adelante, en la Sección 3.4.





**Figura 3.3:** Las líneas rojas verticales indican el ancho de banda utilizado en la Figura 3.4, que corresponde al de la base de datos RRUFF.



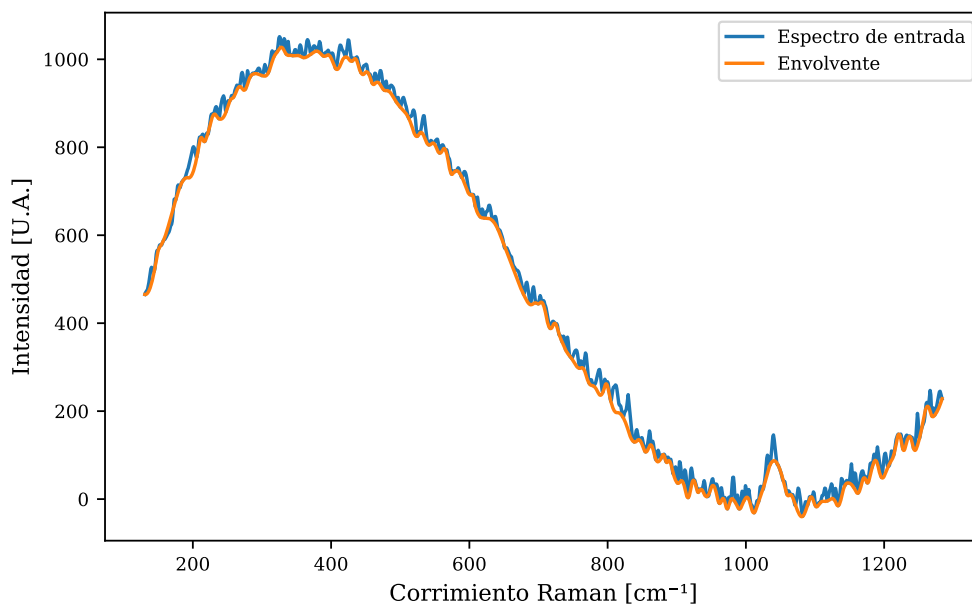
**Figura 3.4:** Corrección de línea base por regresión polinomial al espectro de turquesa truncado a un ancho de banda cercano a los 1200  $[cm^{-1}]$ . El espectro de entrada en todo el ancho de banda se muestra en la Figura 3.3

### 3.3.2. Filtro Savitzky-Golay y señal envolvente

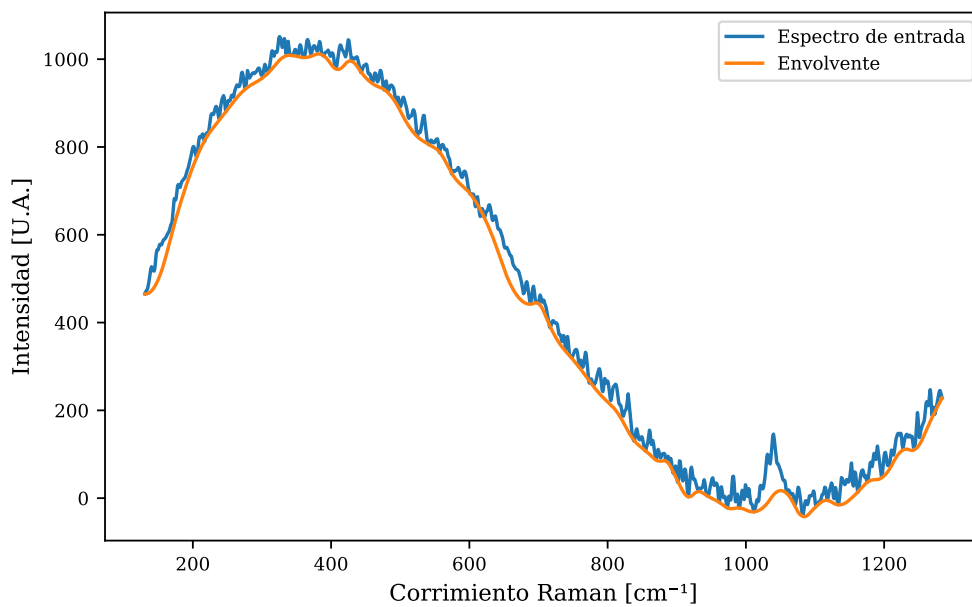
#### Envolvente

Otro procedimiento que se exploró para la corrección de la línea base, es el de calcular una señal envolvente por debajo del espectro de entrada, con el fin de sustraer la envolvente del espectro de entrada y así remover el efecto de fluorescencia.

Para calcular la envolvente se utiliza un pseudocódigo que calcula todos los mínimos locales del espectro de entrada. Una vez calculados los puntos de los mínimos locales, se calcula un modelo que ajuste esos puntos con *splines* cuadráticos. Finalmente, se evalúa el modelo en el dominio requerido para obtener una envolvente por debajo del espectro de entrada. En la Figura 3.5a, se muestra la envolvente para el espectro de una turquesa, el cual no es tan suave como se deseaba. Por esta razón, se decidió aplicar nuevamente el pseudocódigo que calcula la envolvente de la señal, como lo muestra la Figura 3.5b. Se obtiene una envolvente más suave cuando se aplica el pseudocódigo por segunda vez, no obstante se puede mejorar el resultado utilizando el filtro Savitzky-Golay.



(a)



(b)

**Figura 3.5:** Cálculo de envolventes a partir de quadratic splines: (a) utilizando una vez el algoritmo; (b) utilizando el algoritmo sobre la envolvente calculada en (a).

### Filtro Savitzky-Golay

Se utilizará el filtro propuesto por [Savitzky y Golay \(1964\)](#) para lograr el objetivo de obtener una señal suave. En esencia, el método realiza una regresión polinómica local en una serie de datos para determinar el valor suavizado de cada punto mediante un procedimiento de mínimos cuadrados. Se realiza en una ventana móvil de longitud  $2k + 1$  para obtener el valor suavizado mediante mínimos cuadrados

$$x_j^i = a_0 + a_1j + a_2j^2 + \cdots + a_kj^k, \quad (3.10)$$

donde

$$j = -k, -k + 1, \dots, k - 1, k,$$

$i$  está caracterizada por la posición central de la ventana móvil en los datos de entrada. Ahora se introduce la aplicación del método en forma matricial. El valor suavizado en la ventana se define como

$$\begin{aligned} X_i &= Ma \\ X_i &= [x_{-k}^i, x_{-k+1}^i, \dots, x_0^i, \dots, x_{k-1}^i, x_k^i]^T, \\ a &= [a_0, a_1, \dots, a_k]^T, \\ M &= \begin{bmatrix} 1 & -k & (-k)^2 & \cdots & (-k)^k \\ 1 & \cdots & \cdots & \cdots & (-k+1)^k \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & k-1 & (k-1)^2 & \cdots & (k-1)^k \\ 1 & k & k^2 & \cdots & k^k \end{bmatrix}_{(2k+1) \times (k+1)}. \end{aligned} \quad (3.11)$$

Para calcular  $a$ , la Ecuación 3.11 puede reescribirse como

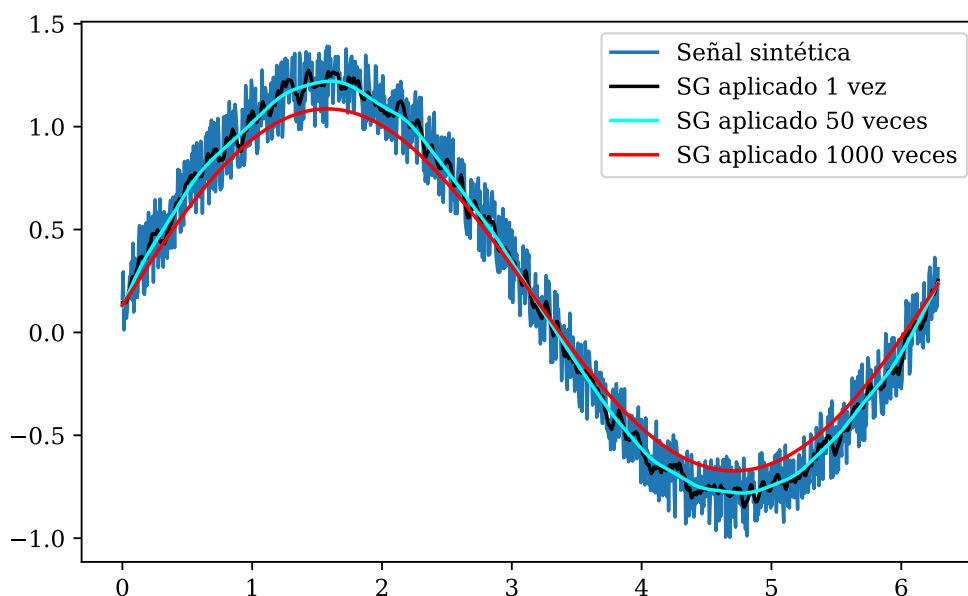
$$M^T Ma = M^T X_i \quad (3.12)$$

Entonces, la solución de mínimos cuadrados de la Ecuación 3.12 es  $\hat{a} = (M^T M)^{-1} M^T X_i$ . Substituyendo  $\hat{a}$  en la Ecuación 3.11 se obtiene el valor suavizado de mínimos cuadrados por la ventana móvil polinomial

$$\begin{aligned} \hat{X}_i &= [\widehat{x_{-k}^i}, \widehat{x_{-k+1}^i}, \dots, \widehat{x_0^i}, \dots, \widehat{x_{k-1}^i}, \widehat{x_k^i}]^T = M\hat{a} \\ &= M (M^T M)^{-1} M^T X_i = GX_i, \end{aligned} \quad (3.13)$$

donde  $G \in \mathbb{R}^{(2k+1) \times (2k+1)}$  es el filtro de Savitzky-Golay,  $X_i$  es el espectro de entrada y  $\hat{X}_i$  es el espectro suavizado.

Se tiene ahora un método para suavizar la envolvente y lograr así una señal suave que permita remover el ruido de fondo y corregir la línea base. Ahora bien, para aprovechar el potencial completo del filtro y suavizar la señal sin perder características importantes del espectro, se realiza el filtro de manera iterativa. La Figura 3.6 muestra la importancia de escoger adecuadamente el número de veces que se aplicará el filtro. Debido a que si se aplica pocas veces, no se suaviza la señal y si se aplica demasiadas, se obtiene una señal cada vez más plana y se pierden características importantes del espectro.



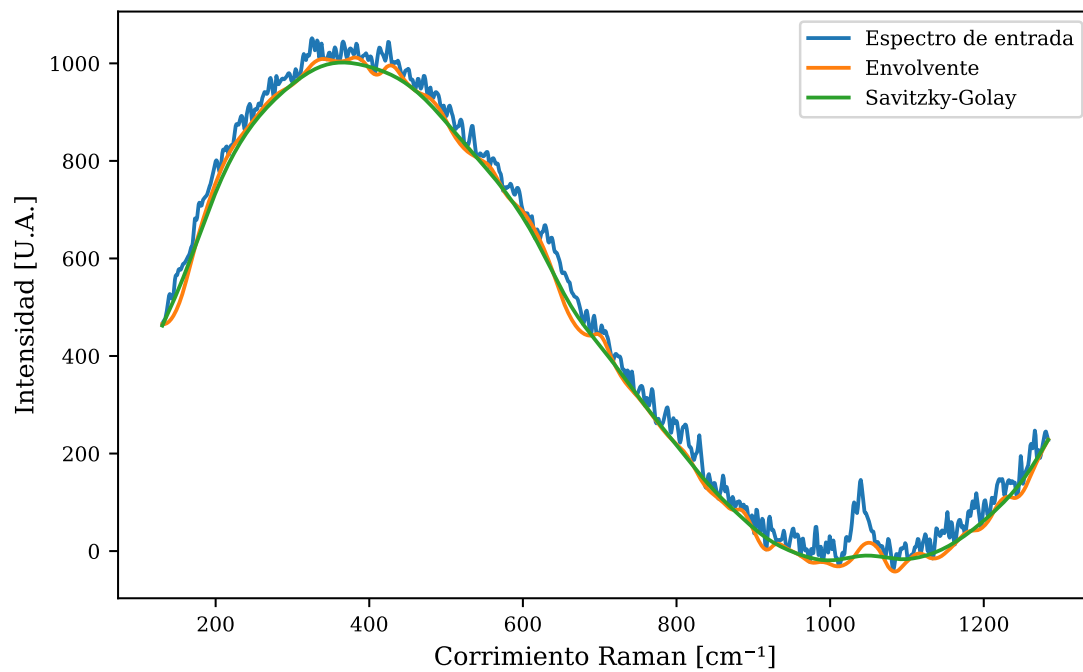
**Figura 3.6:** Diferentes iteraciones del filtro Savitzky y Golay aplicado a una señal sintética.

Es importante señalar que en este trabajo, y para el conjunto de datos con los que se trabajó, se escogió un total de 100 iteraciones con un tamaño de ventana dado por  $(2k + 1)$  donde  $k = 5$ ; para el grado de la aproximación por mínimos cuadrados se escogió una aproximación lineal. En caso de trabajar con muestras tomadas bajo diferentes condiciones (e.g. con un láser de diferente longitud de onda) es de gran importancia escoger de nuevo los parámetros que funcionen mejor para los espectros estudiados.

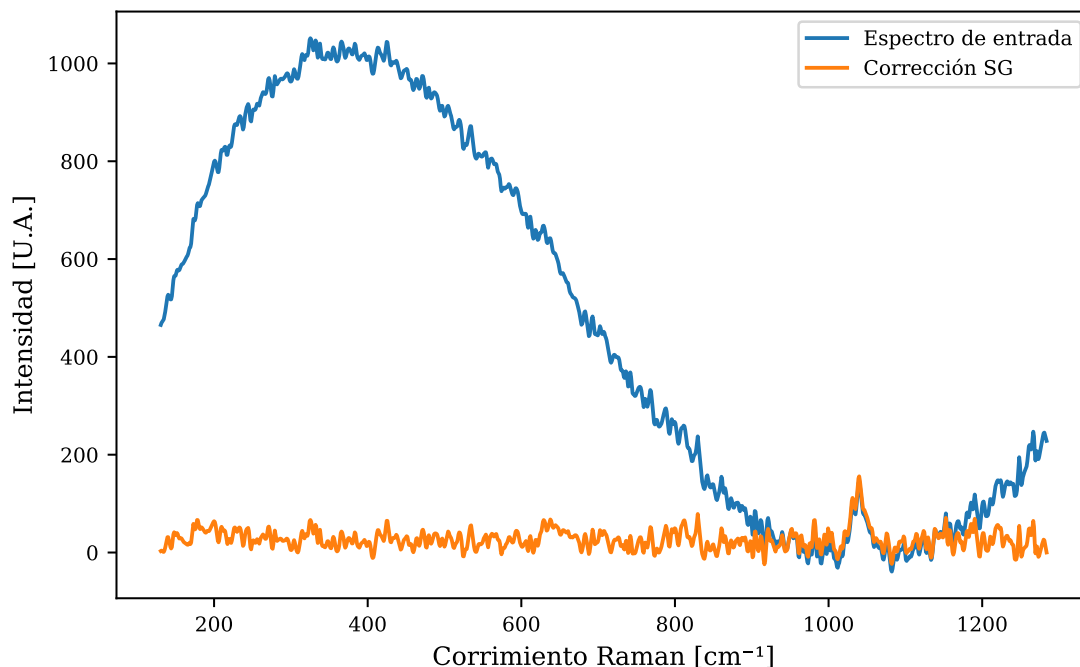
Una vez que se obtiene el ruido de fondo mediante el cálculo de la envolvente y la aplicación del filtro Savitzky-Golay (ver Figura 3.7), se sustrae del espectro inicial para conseguir así, la corrección por línea base. En la Figura 3.8 se ilustra este procedimiento para el espectro de una turquesa.

### 3. METODOLOGÍA PARA LA CORRECCIÓN DE ESPECTROS

---



**Figura 3.7:** Filtro Savitzky Golay aplicado a la envoltente del espectro de turquesa.



**Figura 3.8:** Corrección de línea base por el método de la envolvente y el filtro Savitzky-Golay.

### 3.3.3. Mínimos cuadrados penalizados con ponderación iterativamente adaptable (airPLS)

El último método que se presentará en este trabajo para realizar la corrección de línea base es el propuesto por [Zhang et al. \(2010a\)](#).

#### Pseudocódigo de mínimos cuadrados penalizados

El pseudocódigo de mínimos cuadrados penalizados es un método de suavizado flexible publicado por [Whittaker \(1922\)](#). Después, [Silverman \(1993\)](#) desarrolló una nueva técnica de suavizado en estadística, que se denominó método de penalización de rugosidad. El pseudocódigo de mínimos cuadrados penalizados puede considerarse como una penalización de rugosidad suave por mínimos cuadrados, que equilibra la fidelidad a los datos originales y la rugosidad de los datos ajustados. [Liang et al. \(1999\)](#) lo introdujeron en Química como una técnica de suavizado para mejorar la detección de señales y la resolución de componentes químicos con concentraciones muy bajas en datos cromatográficos.

### 3. METODOLOGÍA PARA LA CORRECCIÓN DE ESPECTROS

---

Suponiendo que  $\mathbf{x}$  es el vector de las señales analíticas, y  $\mathbf{z}$  es el vector ajustado; ambos con longitud  $m$ . La fidelidad de  $\mathbf{z}$  a  $\mathbf{x}$  se puede expresar como la suma de los errores al cuadrado entre ellos,

$$F = \sum_{i=1}^m (x_i - z_i)^2. \quad (3.14)$$

La rugosidad de los datos ajustados  $\mathbf{z}$  se puede escribir como la suma de sus diferencias al cuadrado,

$$R = \sum_{i=2}^m (z_i - z_{i-1})^2 = \sum_{i=1}^{m-1} (\Delta z_i)^2. \quad (3.15)$$

La primera penalización por diferencias se adopta para simplificar la presentación aquí. En la mayoría de los casos, el cuadrado de las segundas diferencias puede ser una forma natural de cuantificar la rugosidad.

El balance entre la fidelidad y la suavidad se puede medir como la fidelidad más penalizaciones en la rugosidad, y puede ser dada por

$$Q = F + \lambda R = \|\mathbf{x} - \mathbf{z}\|^2 + \lambda \|\mathbf{D}\mathbf{z}\|^2. \quad (3.16)$$

Aquí  $\lambda$  puede ser ajustada por el usuario. Una  $\lambda$  más grande se traduce en un vector más suavizado. El balance entre la fidelidad y la suavidad se puede lograr modificando este parámetro.  $\mathbf{D}$  es el operador de diferencias finitas tal que  $\mathbf{D}\mathbf{z} = \Delta\mathbf{z}$ .

Al calcular el vector de derivadas parciales e igualarlo a cero ( $\frac{\partial Q}{\partial \mathbf{z}} = 0$ ), se obtiene el sistema lineal de ecuaciones:

$$(\mathbf{I} + \lambda\mathbf{D}'\mathbf{D})\mathbf{z} = \mathbf{x}. \quad (3.17)$$

La Ecuación 3.17 es un método de suavizado usando el pseudocódigo de mínimos cuadrados penalizados. Para corregir la línea base utilizando el pseudocódigo de mínimos cuadrados penalizados, [Zhang et al. \(2010b\)](#) introdujo un vector de peso de fidelidad, y estableció a cero el vector de pesos en una posición correspondiente a los picos de  $\mathbf{x}$ . La fidelidad de  $\mathbf{z}$  a  $\mathbf{x}$  cambia entonces a

$$F = \sum_{i=1}^m w_i (x_i - z_i)^2 = (\mathbf{x} - \mathbf{z})' \mathbf{W} (\mathbf{x} - \mathbf{z}), \quad (3.18)$$

donde  $\mathbf{W}$  es una matriz diagonal con  $w_i$  en su diagonal. La Ecuación 3.17 cambia a



$$(\mathbf{W} + \lambda \mathbf{D}'\mathbf{D}) \mathbf{z} = \mathbf{W}\mathbf{x}. \quad (3.19)$$

Resolviendo las ecuaciones lineales anteriores, el vector ajustado se puede obtener mediante

$$\mathbf{z} = (\mathbf{W} + \lambda \mathbf{D}'\mathbf{D})^{-1} \mathbf{W}\mathbf{x}. \quad (3.20)$$

La corrección de línea base de Zhang et al. necesita detección de picos en la señal y aunque el pseudocódigo es correcto, consume mucho tiempo. Es por eso que se propone el pseudocódigo de ponderación iterativa adaptativa, para reemplazar la detección de picos y pasos de tratamiento adicional.

### Procedimiento de ponderación iterativamente adaptable

El procedimiento de ponderación iterativamente adaptable es similar al de mínimos cuadrados pesados y al de mínimos cuadrados repesados iterativamente, pero utiliza diferentes cálculos de los pesos y añade un penalizador para controlar la suavidad de la línea base ajustada. Cada paso del pseudocódigo involucra resolver un problema por mínimos cuadrados penalizados pesados de la forma

$$Q^t = \sum_{i=1}^m w_i^t |x_i - z_i^t|^2 + \lambda \sum_{j=2}^m |z_j^t - z_{j-1}^t|^2. \quad (3.21)$$

El vector de peso  $\mathbf{w}$  se obtiene usando un método iterativo. Se debe proporcionar un valor inicial  $w^0 = 1$  para el paso inicial. Después de la inicialización, el vector  $\mathbf{w}$  en cada paso iterativo  $t$ , puede ser obtenido usando la siguiente expresión

$$w_i^t = \begin{cases} 0 & x_i \geq z_i^{t-1} \\ e^{\frac{t(x_i - z_i^{t-1})}{|a^t|}} & x_i < z_i^{t-1} \end{cases}, \quad (3.22)$$

donde  $\mathbf{d}^t$  es un vector con los elementos negativos de las diferencias entre  $\mathbf{x}$  y  $\mathbf{z}^{t-1}$  en la iteración  $t$ .

El valor ajustado  $\mathbf{z}^{t-1}$  en la iteración previa ( $t-1$ ) es un candidato a línea base. Si el valor del  $i$ -ésimo punto es mayor que el candidato de la línea base, se puede considerar como parte del pico. Entonces su peso se fija en cero para ignorarlo en la siguiente iteración de ajuste. En el pseudocódigo airPLS, los métodos repesados e iterativos son usados para eliminar los puntos de los picos gradualmente y automáticamente y preservar los puntos de la línea base en el vector de pesos  $\mathbf{w}$ .

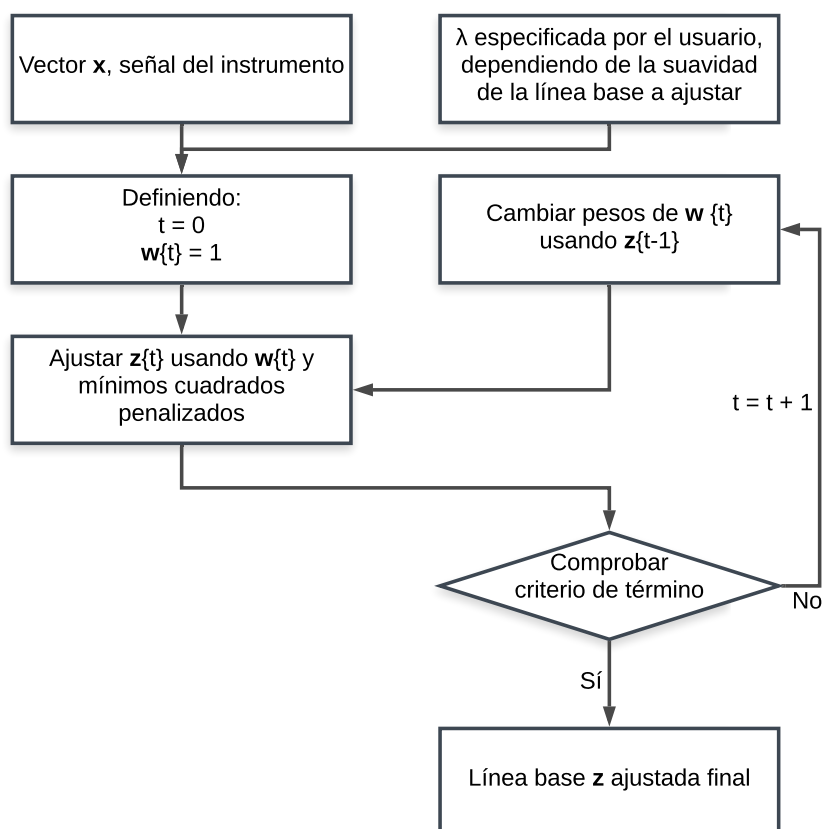
La iteración se detendrá ya sea con el número máximo de iteraciones o cuando el criterio de término se cumpla. El criterio de término se define como

### 3. METODOLOGÍA PARA LA CORRECCIÓN DE ESPECTROS

---

$$|\mathbf{d}_t| < 0.001 \times |\mathbf{x}|. \quad (3.23)$$

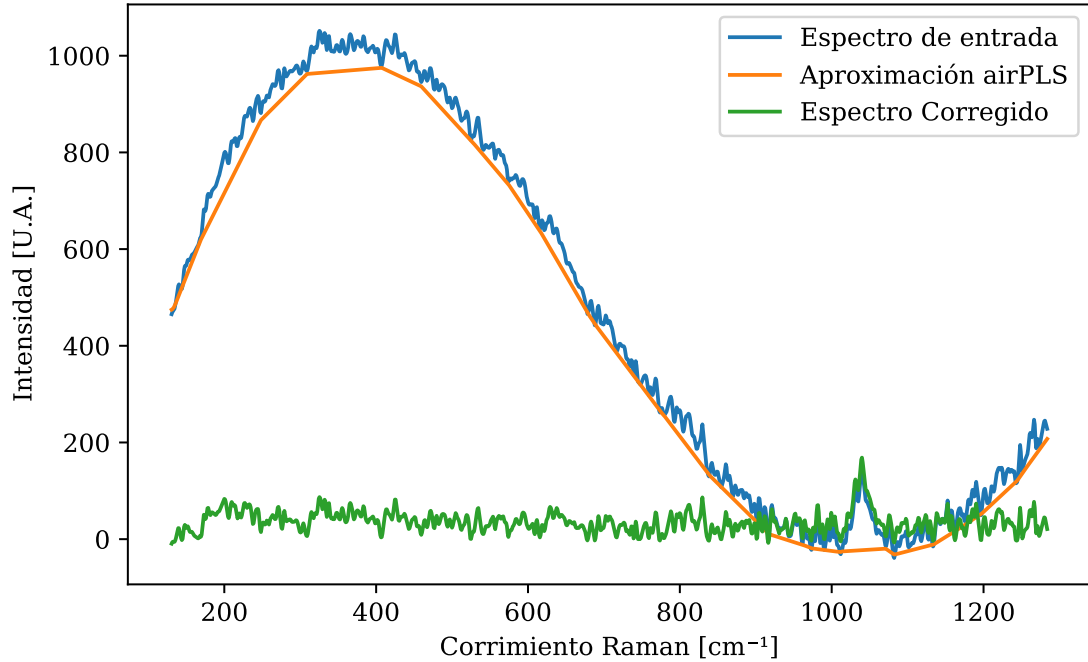
El flujo del pseudocódigo airPLS puede observarse de manera esquemática en la Figura 3.9



**Figura 3.9:** Diagrama de flujo describiendo el esquema de trabajo del pseudocódigo airPLS.

Los resultados del método airPLS se muestran en la Figura 3.10. Después de probar con distintas configuraciones, se llegó a la conclusión de que los parámetros que, generalmente, arrojan los mejores resultados son:

- $\lambda = 10$
- Número máximo de iteraciones = 5
- Grado del polinomio = 1



**Figura 3.10:** Corrección de línea base a un espectro de turquesa por el método de mínimos cuadrados penalizados con ponderación iterativamente adaptable (airPLS).

### 3.4. Filtrado en altas frecuencias

Como se comentó anteriormente y como se muestra en la Figura 3.2, el espectro recolectado se compone, además del efecto de la fotoluminiscencia, de ruido blanco en alta frecuencia.

A pesar de que el espectro corregido por línea base puede conseguir una mejor coincidencia en la base de datos, creemos que este resultado se puede mejorar significativamente atacando el problema del ruido en altas frecuencias. El ruido Gaussiano es inherente a las condiciones bajo las que se recolectan los espectros.

Para remover el ruido en altas frecuencias se escogió un filtro pasa-bajas tipo Butterworth. Después de probar con diferentes parámetros para el conjunto de datos trabajado en esta tesis, se llegó a la elección de los siguientes parámetros:

- Frecuencia de muestreo: 1000
- Frecuencia de corte: 50
- Orden del filtro: 9

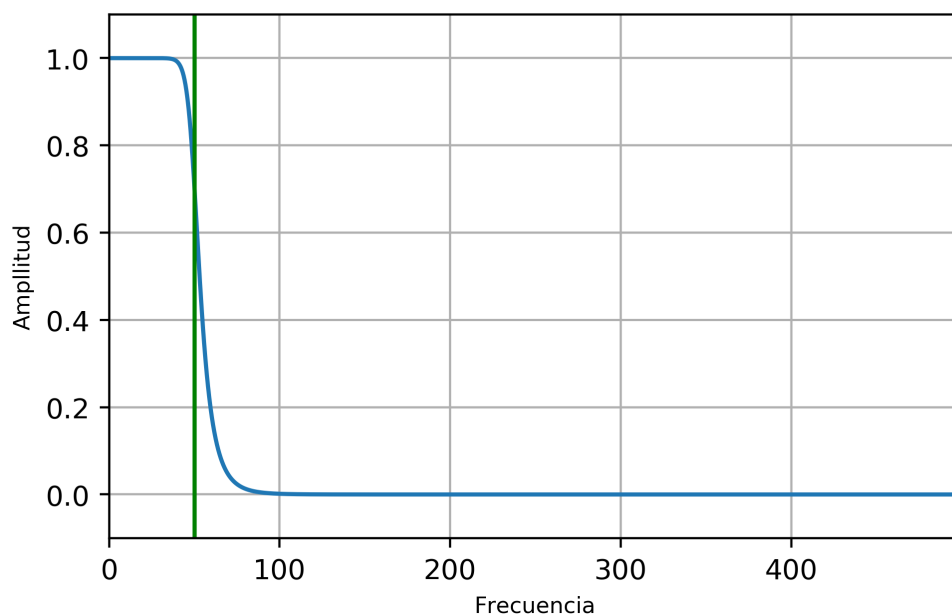
### 3. METODOLOGÍA PARA LA CORRECCIÓN DE ESPECTROS

---

La elección de estos parámetros se basó en la búsqueda de filtrar el ruido de alta frecuencia, afectando lo menos posible la amplitud de los picos en frecuencias características.

De nuevo es importante mencionar que los parámetros del filtro deben ser actualizados al momento de trabajar con un conjunto de datos diferente.

La respuesta del filtro utilizado puede apreciarse en la Figura 3.11.



**Figura 3.11:** Respuesta en frecuencia del filtro Butterworth pasa-bajas. En azul la amplitud de la señal y en verde la frecuencia de corte del filtro Butterworth.

### 3.5. Efecto de amplitudes negativas

Al sustraer la línea base para la corrección de la fluorescencia, se obtienen algunos valores negativos de amplitud. Para corregir este efecto se aplica la ecuación

$$x_c = x - \min(x), \quad (3.24)$$

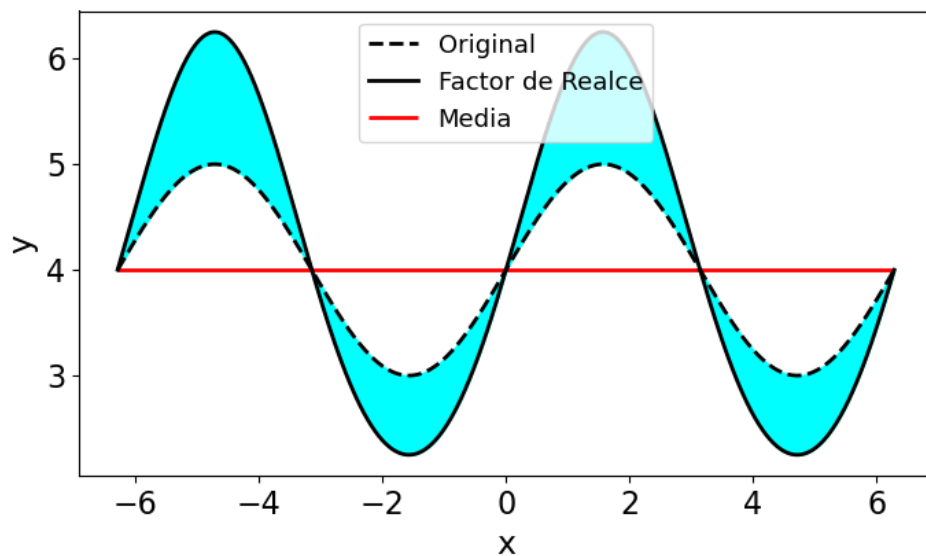
donde  $x_c$  es el espectro sin valores negativos,  $x$  es el espectro filtrado y con la corrección por línea base y  $\min(x)$  es el valor mínimo global de  $x$ .

### 3.6. Factor de realce

El objetivo de la Espectroscopía Raman es determinar una especie mineral mediante su estructura, la cual va a estar en parte determinada por su composición. La amplitud de la señal en una determinada frecuencia es proporcional a la concentración del elemento químico asociado a esa frecuencia característica. Lamentablemente, parte de la amplitud de los picos de los espectros se compromete al remover el ruido de la fluorescencia con los algoritmos detallados en este trabajo. Con el fin de realzar la amplitud de la señal en las frecuencias que caracterizan a la muestra, proponemos un factor que modifica la amplitud de la señal. Este factor atenúa los valores del espectro que se encuentran por debajo de la media y amplifica los que se encuentran por encima de la misma. Los valores que se encuentran sobre la media no sufren modificaciones. El factor se aplica según la expresión

$$y_c = y \underbrace{\left(1 - \frac{\text{promedio} - y}{\text{promedio}}\right)}_{\text{Factor de realce}}, \quad (3.25)$$

donde *promedio* es el valor promedio de todo el espectro y  $y_c$  representa el espectro con el efecto del factor de realce. El efecto esencial del factor sobre una señal digital se puede observar en la Figura 3.12.



**Figura 3.12:** Ejemplo del efecto del factor de realce sobre una señal.

## 3.7. Normalización espectral

Antes de poder calcular el coeficiente de correlación, es conveniente que las señales que van a compararse se encuentren normalizadas. Para ello se realiza

$$\begin{aligned} y_n &= \frac{y}{\max(y)} \\ yR_n &= \frac{yR}{\max(yR)}. \end{aligned} \tag{3.26}$$

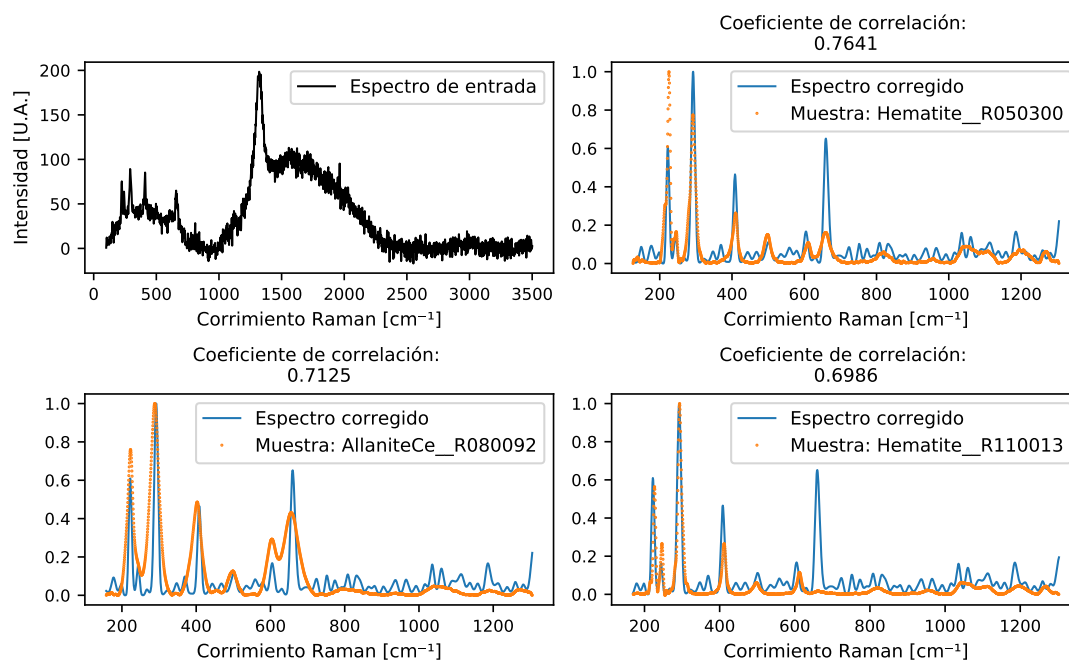
Donde  $y_n$  y  $yR_n$  son los espectros normalizados,  $\max(y)$  y  $\max(yR)$  son los valores máximos globales para los espectros sin normalizar, y  $y$  y  $yR$  son los espectros sin normalizar ( $yR$  es el espectro de RRUFF).

## 3.8. Correlación espectral

Los espectros que se recolectaron por parte del proyecto del INAH y los espectros de RRUFF no comparten el mismo muestreo. La señal de las muestras facilitadas por el INAH contienen un dato cada  $0.964[cm^{-1}]$ , mientras que los espectros de RRUFF presentan un dato cada  $0.482[cm^{-1}]$ . Con el fin de realizar una correlación con la misma densidad de datos, se realizará una interpolación 1D. Como dominio común, se escogió el dominio de los espectros del IGUM ( $x$ ), por ser los que presentan una menor densidad de datos. Una vez que se tienen ambos espectros con el mismo muestreo, se procede al cálculo del coeficiente de correlación con la ayuda de la función *stats.pearsonr* de la paquetería *scipy*.

Debido a que todo el proceso anterior se aplica para cada espectro de la base de datos, al final del proceso se obtienen tantos coeficientes de correlación como espectros contenga la base de datos. Es por esto que se almacenan los coeficientes en un vector del tamaño de la base de datos. Con la excepción del método de corrección de línea base por el método polinomial, en este caso se tiene una matriz de  $(n \times m)$  donde  $n$  es el tamaño de la base de datos y  $m$  es el número de órdenes de polinomios que se aplican para la corrección.

Con el fin de presentar los resultados, se grafican el espectro de entrada y los espectros de RRUFF correspondientes a los tres coeficientes de correlación más altos (ver Figura 3.13). En ocasiones, la figura con los tres espectros puede no ser concluyente, por lo que también se imprimen en terminal, los diez coeficientes de correlación más altos, así como el nombre de la muestra que les corresponde.



**Figura 3.13:** Resultado de aplicar la metodología propuesta con airPLS a un espectro de Hematita.

### 3.9. Pseudocódigos

En esta sección se presentan los pseudocódigos para el procesamiento de espectros Raman de cada uno de los métodos de corrección de la línea base.

### 3. METODOLOGÍA PARA LA CORRECCIÓN DE ESPECTROS

---

**Parámetros:** Grado mínimo y máximo para el ajuste polinomial,  $g_{ini}$  y  $g_{max}$ .

**Datos:** Espectro a analizar,  $[w_0, s_0]$ , y la base de datos con los espectros de referencia de RRUFF,  $[w_i, s_i] \forall i \in \{1, \dots, N\}$ .

**Resultado:** Máximos coeficientes de correlación entre el espectro a analizar y espectros de RRUFF.

- 1 Se inicializa el orden del polinomio,  $g_i = g_{ini}$ .
- 2 **mientras**  $g_i \leq g_{max}$  **hacer**
- 3     **para cada** espectro  $[w_i, s_i]$  **de la base de datos** **hacer**
  - Ajustar el dominio del número de onda  $w_0$  de acuerdo a  $w_i$ .
  - Ajustar la línea base:  
 $s_0 = s_0 -$  la aproximación polinomial de  $s_0$  de orden  $g_i$ .
  - Aplicar un filtro Butterworth pasa-bajas a  $s_0$  y  $s_i$ .
  - Eliminar valores negativos:  
 $s_0 = s_0 - \min(s_0)$ .
  - Aplicar el factor de realce a  $s_0$  (ver Sección 3.6).
  - Normalizar los espectros  $s_0$  y  $s_i$
  - $cc(g_i, i) =$  coeficiente de correlación entre  $s_i$  y  $s_0$
- 4     Incrementar el orden polinomial,  $g_i = g_i + 1$
- 5 Se extraen los tres coeficientes de correlación mayores de  $cc$ ;

**Pseudocódigo 1:** Procesamiento de espectros de Raman para corrección de la línea base por el método de regresión polinomial



**Datos:** Espectro a analizar,  $[w_0, s_0]$ , y la base de datos con los espectros de referencia de RRUFF,  $[w_i, s_i] \forall i \in \{1, \dots, N\}$ .

**Resultado:** Máximos coeficientes de correlación entre el espectro a analizar y espectros de RRUFF.

**1 para cada espectro  $[w_i, s_i]$  de la base de datos hacer**

- Ajustar el dominio del número de onda  $w_0$  de acuerdo a  $w_i$ .
- Calcular la envolvente de  $s_i$ .  $env_1$
- Aplicar un filtro Savitzky-Golay a la envolvente de  $env_1$ .  $s_c$
- Ajuste de la línea base:  
 $s_0 = s_0 - s_c$
- Aplicar un filtro Butterworth pasa-bajas a  $s_0$  y  $s_i$ .
- Eliminar valores negativos:  
 $s_0 = s_0 - \min(s_0)$ .
- Se aplica el factor de realce a  $s_0$  (ver Sección 3.6).
- Se normalizan los espectros  $s_0$  y  $s_i$ .
- $cc(g_i) =$  coeficiente de correlación entre  $s_i$  y  $s_0$ .

**2** Se extraen los tres coeficientes de correlación mayores de  $cc$

**Pseudocódigo 2:** Procesamiento de espectros de Raman para corrección de la línea base por el método del filtro Savitzky-Golay.

### 3. METODOLOGÍA PARA LA CORRECCIÓN DE ESPECTROS

---

**Datos:** Espectro a analizar,  $[w_0, s_0]$ , y la base de datos con los espectros de referencia de RRUFF,  $[w_i, s_i] \forall i \in \{1, \dots, N\}$ .

**Resultado:** Máximos coeficientes de correlación entre el espectro a analizar y espectros de RRUFF.

1 para cada espectro  $[w_i, s_i]$  de la base de datos hacer

- Ajustar el dominio del número de onda  $w_0$  de acuerdo a  $w_i$ .
- Calcular el ajuste de  $s_0$  por medio de airPLS.  $s_c$
- Ajuste de la línea base:  
 $s_0 = s_0 - s_c$
- Aplicar un filtro Butterworth pasa-bajas a  $s_0$  y  $s_i$ .
- Eliminar valores negativos:  
 $s_0 = s_0 - \min(s_0)$ .
- Aplicar el factor de realce a  $s_0$  (ver Sección 3.6).
- Normalizar los espectros  $s_0$  y  $s_i$ .
- $cc(g_i) =$  coeficiente de correlación entre  $s_i$  y  $s_0$ .

2 Se extraen los tres coeficientes de correlación mayores de  $cc$

**Pseudocódigo 3:** Procesamiento de espectros de Raman para corrección de la línea base por el método de airPLS.

---

## Capítulo 4

# Validación de la metodología

---

<b>4.1. Espectros de referencia . . . . .</b>	<b>35</b>
<b>4.2. Regresión polinomial . . . . .</b>	<b>36</b>
<b>4.3. Filtro Savitzky-Golay y señal envolvente . . . . .</b>	<b>41</b>
<b>4.4. Mínimos Cuadrados Penalizados con Ponderación Iterativamente Adaptable (airPLS) . . . . .</b>	<b>44</b>
<b>4.5. Comparación entre las metodologías . . . . .</b>	<b>47</b>

---

En este capítulo, se expone la eficacia y los resultados de la metodología propuesta para la corrección espectral. Se aplican los tres métodos propuestos para corregir la línea base, a tres espectros que presentan ruido asociado con fluorescencia.

### 4.1. Espectros de referencia

Los espectros que se escogieron para presentar los resultados de los métodos propuestos, se recolectaron gracias al apoyo del IGUM, al facilitar el acceso al Laboratorio de Microanálisis del CeMIEGeo. La toma de las muestras fue realizada los días 3 y 4 de mayo de 2019.

Los espectros analizados en esta sección, los escogimos por ser muestras que se clasificaron con la ayuda de otro método espectroscópico. La clasificación por espectroscopia de rayos X de dispersión de energía (EDS), fue posible gracias a la Microsonda Electrónica del Laboratorio de Microanálisis. La Microsonda Electrónica realiza análisis químicos cuantitativos (no destructivos) a escalas mi-

## 4. VALIDACIÓN DE LA METODOLOGÍA

---

crométricas empleando los rayos X que se producen por la interacción de un haz de electrones con la muestra.

Los espectros a analizar pertenecen a muestras de los siguientes minerales:

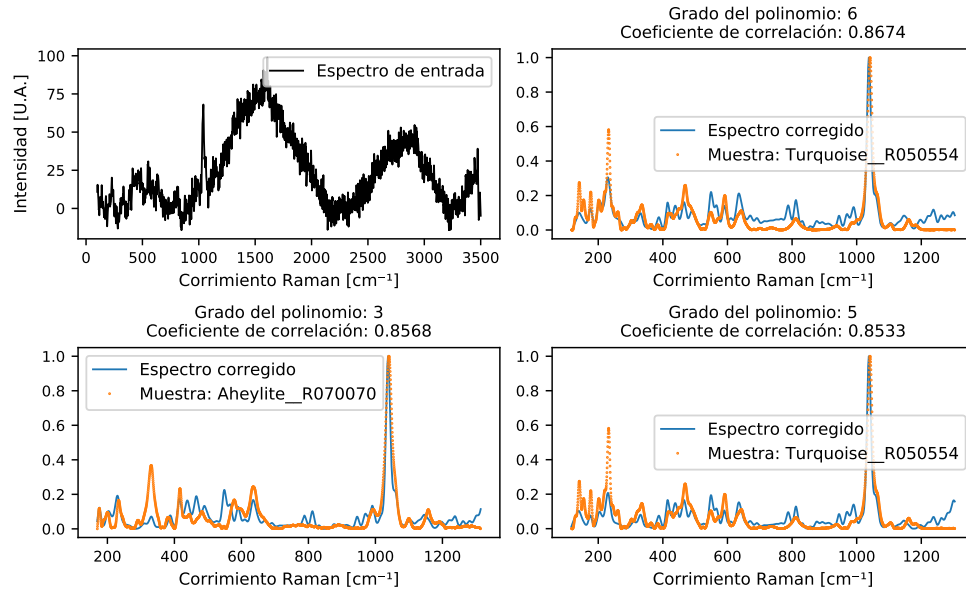
- |             |   |
|-------------|---|
| 1. Turquesa | $\text{CuAl}_6(\text{PO}_4)_4(\text{OH})_8 \cdot 4\text{H}_2\text{O}$ |
| 2. Jarosita | $\text{KFe}^{3+}_3(\text{SO}_4)_2(\text{OH})_6$                       |
| 3. Dickita  | $(\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4)$                     |

### 4.2. Regresión polinomial

Para el caso del ajuste polinomial, se utilizan funciones polinómicas de diferentes grados para ajustar la curva de desviación de la línea base, desde el uno hasta el nueve. Esto, con la finalidad de aproximar, con distintos grados de suavidad, el ruido de la fluorescencia para espectros que presentan diferentes niveles de alteración en su línea base.

#### Turquesa

Comenzando por el espectro de la Turquesa, se puede observar en la Figura [4.1](#) el resultado del código de regresión polinomial.

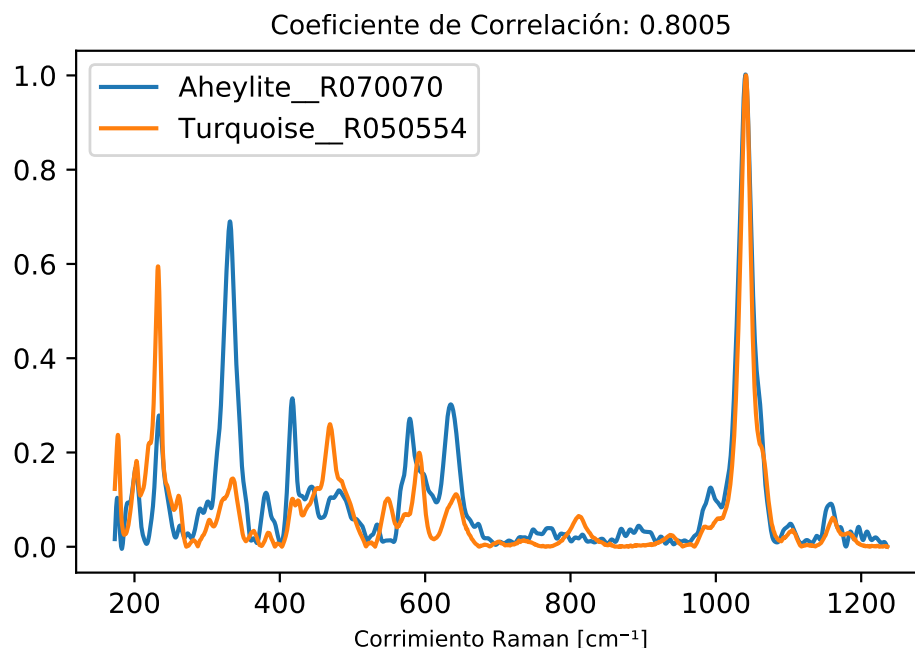


**Figura 4.1:** Resultado de aplicar la metodología de regresión polinomial al espectro de una muestra de turquesa.

En la Figura 4.1 se puede observar que el espectro que obtiene un coeficiente de correlación mayor, es el de una turquesa. Los siguientes dos espectros corresponden a una aheylita ( $\text{Fe}^{2+}\text{Al}_6(\text{PO}_4)_4(\text{OH})_8 \cdot 4\text{H}_2\text{O}$ ) y a una turquesa, respectivamente. Aunque la muestra no es una aheylita, en la Figura 4.2 se puede apreciar que los espectros de turquesa y de aheylita son muy parecidos. La semejanza de los espectros de turquesa y de aheylita son resultado de la similitud en composición y arreglo cristalino. La aheylita pertenece al grupo de la turquesa y ambos minerales poseen un arreglo cristalino triclinico. La similitud que presentan los espectros de la aheylita y la turquesa se aprecian en la Figura 4.2.

#### 4. VALIDACIÓN DE LA METODOLOGÍA

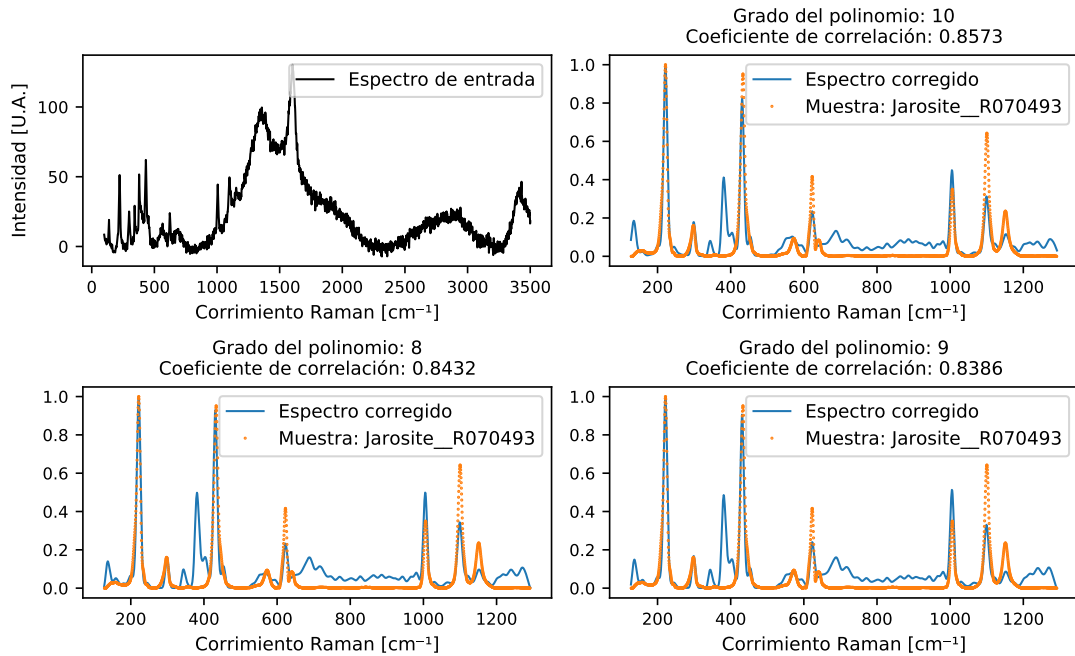
---



**Figura 4.2:** Correlación entre los espectros de referencia de una aheylita y de una turquesa.

### Jarosita

El siguiente espectro a ser analizado es el correspondiente a una muestra de jarosita. Para este caso, los resultados de los tres coeficientes de correlación más altos son los ideales, como se observa en la Figura 4.3.

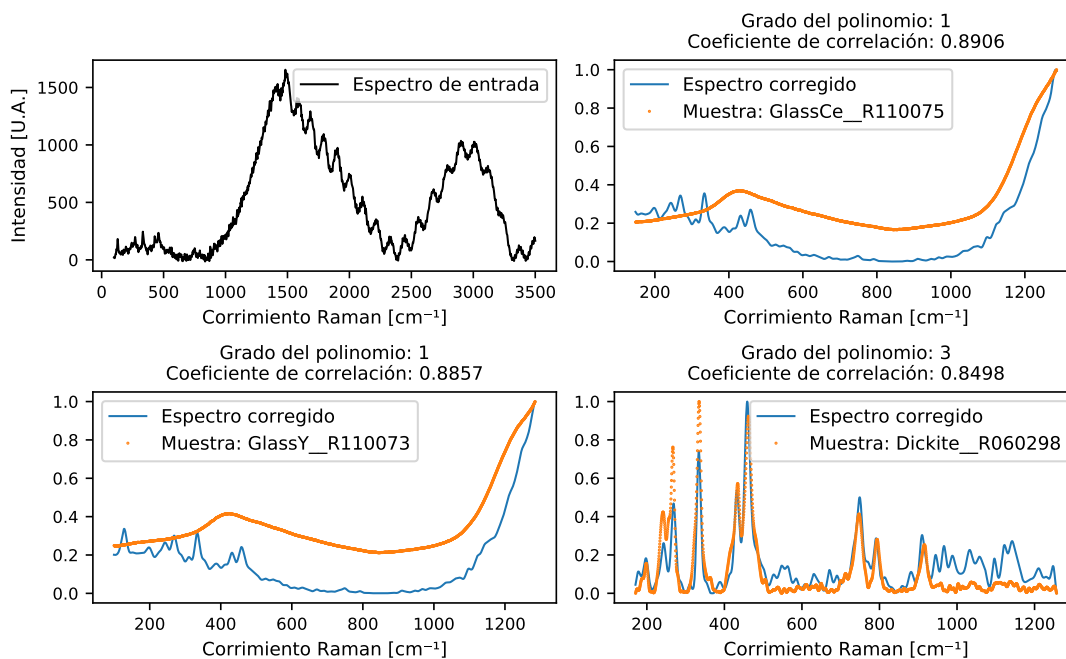


**Figura 4.3:** Resultado de aplicar la metodología de regresión polinomial al espectro de una muestra de jarosita.

## Dickita

El último espectro a ser analizado es el de una muestra de dickita. Los resultados de este espectro en particular, son deficientes. En la Figura 4.4 se observa que el único espectro que cumple con lo esperado es el correspondiente al tercer coeficiente de correlación más alto (abajo derecha). A pesar de que este resultado no es ideal, se puede descartar fácilmente a los primeros dos candidatos debido a que claramente la tendencia no se removió correctamente. Si bien, con la Figura 4.4 no es posible llegar a una clasificación conclusiva del espectro de entrada, se tienen los diez coeficiente de correlación más altos de todo el procesamiento como material adicional para tomar una decisión. Ahora, observando la Figura 4.4 y la Tabla 4.1 se vuelve evidente que, exceptuando los primeros dos espectros, la metodología de regresión polinomial obtiene el resultado correcto.

#### 4. VALIDACIÓN DE LA METODOLOGÍA



**Figura 4.4:** Resultado de aplicar la metodología de regresión polinomial al espectro de una muestra de dickita.

	Muestra	Coefficiente de Correlación
1.	GlassCe__R110075	0.8906
2.	GlassY__R110073	0.8857
3.	Dickite__R060298	0.8498
4.	Dickite__R060298	0.8481
5.	Dickite__R060298	0.8473
6.	Dickite__R060298	0.8452
7.	Dickite__R060298	0.8449
8.	Dickite__R060298	0.8357
9.	Dickite__R060298	0.8339
10.	Dickite__R060298	0.8336

**Tabla 4.1:** Coeficientes de correlación más altos como resultado de la metodología de regresión polinomial al espectro de una dickita.

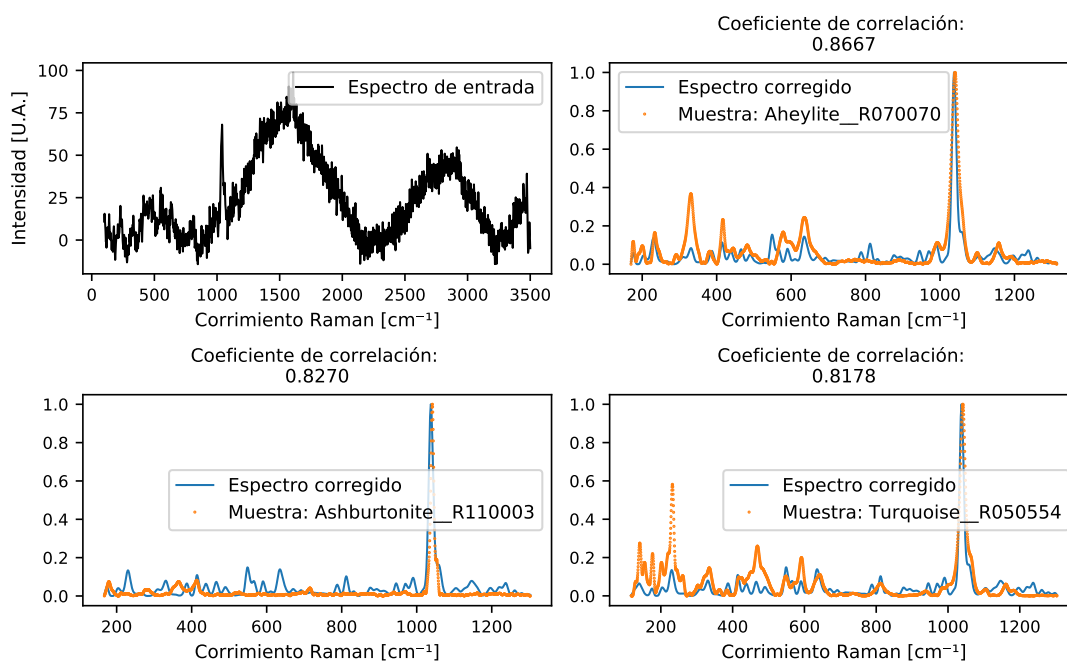


### 4.3. Filtro Savitzky-Golay y señal envolvente

Se seguirá el mismo orden de análisis de los espectros que en la Sección 4.2, esta vez utilizando el método de la envolvente y el filtro Savitzky-Golay para la corrección de la desviación de la línea base ocasionada por la fluorescencia.

#### Turquesa

Comenzando con la turquesa, se muestran los resultados en la Figura 4.5. Los resultados de la Figura 4.5 no sugieren una muestra de turquesa en primera instancia. Al analizar la tabla 4.2, y a pesar de que la turquesa aparece en tres ocasiones entre los diez coeficientes de correlación más altos, el análisis no resulta conclusivo para esta muestra.



**Figura 4.5:** Resultado de aplicar la metodología de la envolvente y el filtro Savitzky-Golay al espectro de una muestra de turquesa.

#### 4. VALIDACIÓN DE LA METODOLOGÍA

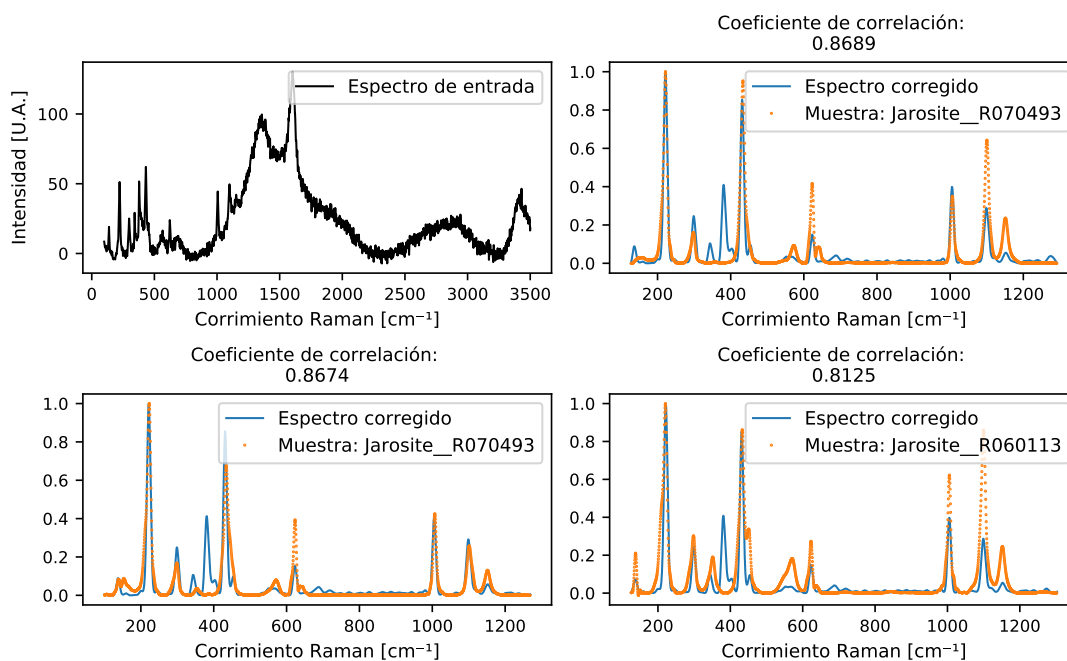
---

	<b>Muestra</b>	<b>Coefficientes de Correlación</b>
<b>1.</b>	Aheylite__R070070	0.8667
<b>2.</b>	Ashburtonite__R110003	0.8270
<b>3.</b>	Turquoise__R050554	0.8178
<b>4.</b>	Turquoise__R050225	0.8050
<b>5.</b>	Gerhardtite__R120082	0.8000
<b>6.</b>	Planerite__R080123	0.7885
<b>7.</b>	Molybdophyllite__R070222	0.7823
<b>8.</b>	Aheylite__R070070	0.7686
<b>9.</b>	Turquoise__R050225	0.7643
<b>10.</b>	Molybdophyllite__R070222	0.6982

**Tabla 4.2:** Coeficientes de correlación más altos como resultado de la metodología de la envolvente y el filtro Savitzky-Golay al espectro de una turquesa.

### Jarosita

Continuando la Jarosita, se presentan los resultados en la Figura 4.6. El espectro de la jarosita analizado con la metodología de Savitzky-Golay, produce resultados ideales; debido a que los tres coeficientes de correlación más altos, corresponden a muestras de jarosita en la base de datos.

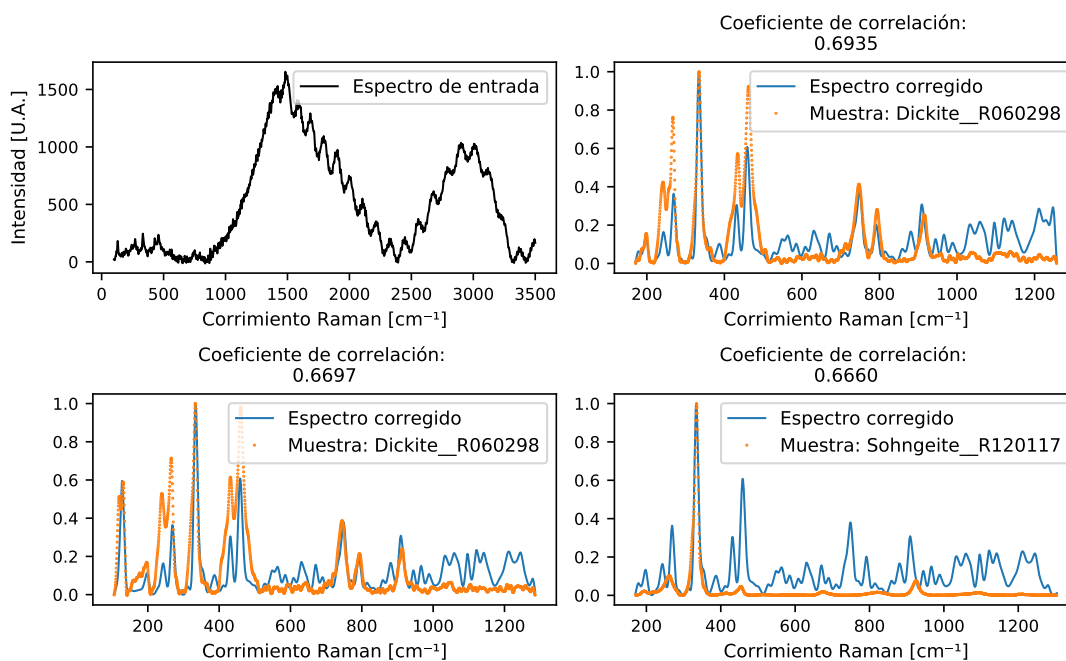


**Figura 4.6:** Resultado de aplicar la metodología de la envolvente y el filtro Savitzky-Golay al espectro de una muestra de jarosita.

## Dickita

Por último, se procede al análisis de la muestra de dickita. Como se observa en la Figura 4.7, los primeros dos espectros sugieren que se trata de una dickita, mas no así el tercer espectro. Una consideración importante para este caso, es que al examinar las primeras dos muestras (arriba derecha y abajo izquierda de la Figura 4.7), se observa una clara correlación con la mayoría de los “picos” entre el espectro corregido y los espectros de la base de datos. En cambio, el espectro de la muestra de sohngeita únicamente correlaciona con un “pico” del espectro corregido.

## 4. VALIDACIÓN DE LA METODOLOGÍA



**Figura 4.7:** Resultado de aplicar la metodología de la envolvente y el filtro Savitzky-Golay al espectro de una muestra de dickita.

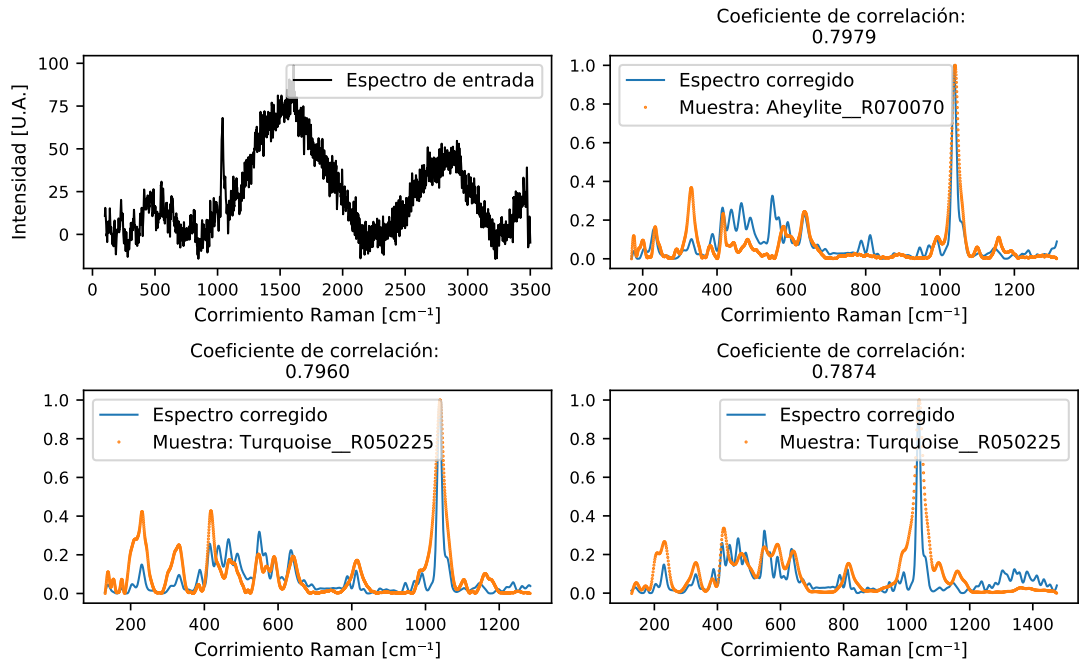
### 4.4. Mínimos Cuadrados Penalizados con Ponderación Iterativamente Adaptable (airPLS)

En esta sección se analizan los espectros con el algoritmo airPLS propuesto por [Zhang et al. \(2010a\)](#).

#### Turquesa

Los resultados al analizar el espectro de una turquesa con el algoritmo airPLS se presentan en la Figura 4.8. Como resultado, se obtienen una aheylita, como el mayor coeficiente de correlación, seguida de dos turquesas. Como se ha mencionado anteriormente, esto es debido a la similitud que presentan los espectros de la aheylita y la turquesa (ver Figura 4.2). Considerando lo anterior, los resultados son aceptables para este espectro.

#### 4.4 Mínimos Cuadrados Penalizados con Ponderación Iterativamente Adaptable (airPLS)

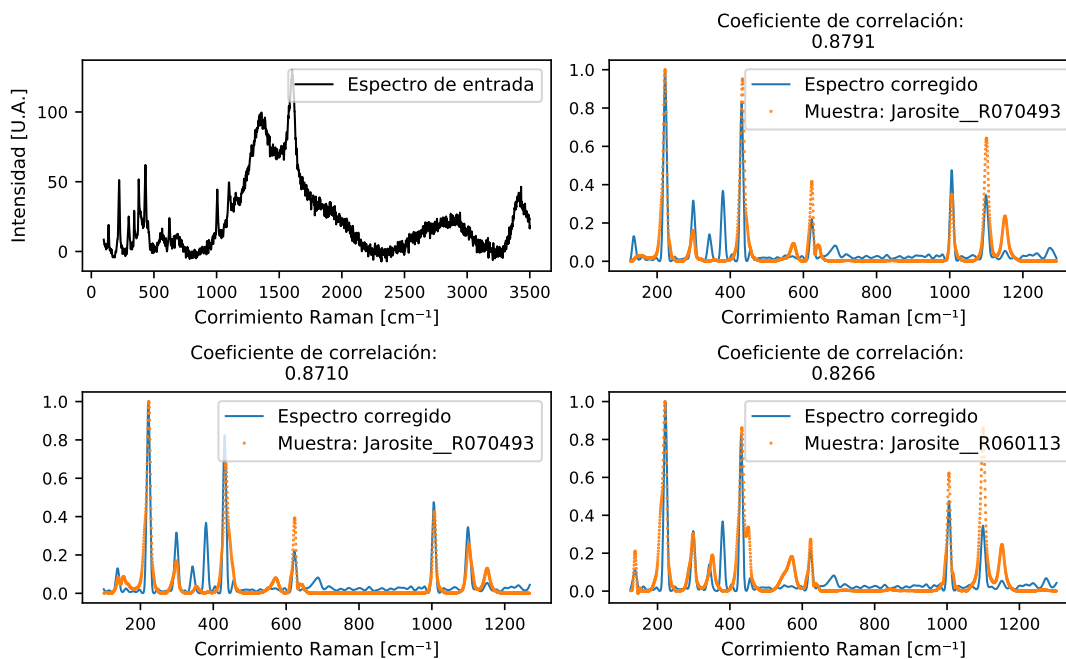


**Figura 4.8:** Resultado de aplicar la metodología airPLS al espectro de una muestra de turquesa.

### Jarosita

La metodología airPLS aplicada al espectro ruidoso de una jarosita genera los resultados presentados en la Figura 4.9. De la Figura 4.9, se observa que el espectro corregido presenta resultados favorables, al coincidir con tres espectros de jarosita de la base de datos de RRUFF.

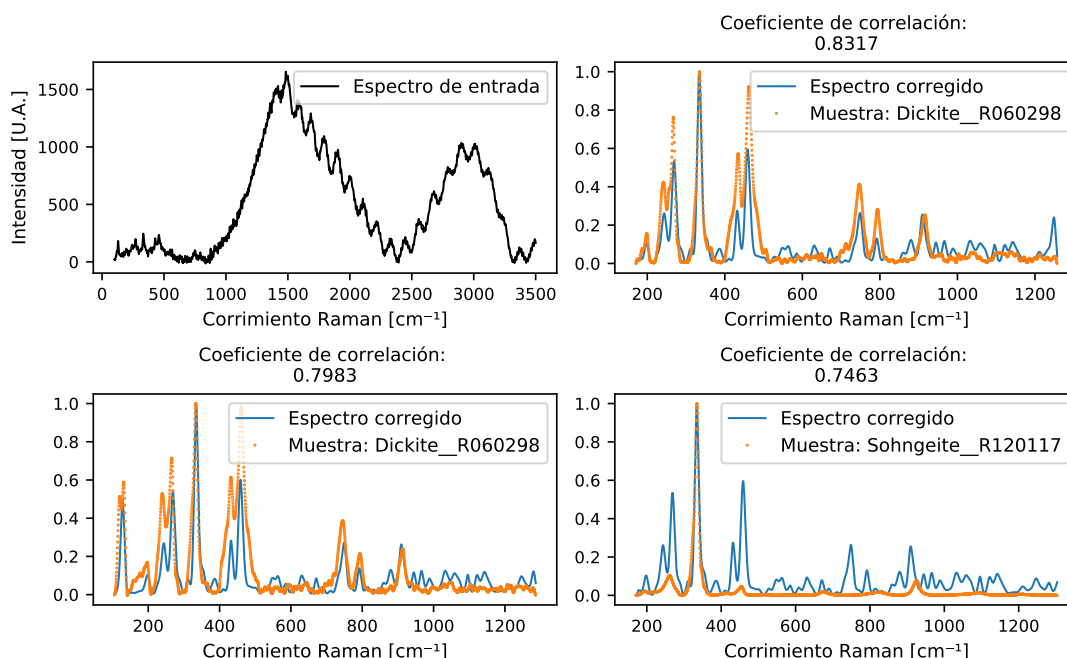
## 4. VALIDACIÓN DE LA METODOLOGÍA



**Figura 4.9:** Resultado de aplicar la metodología airPLS al espectro de una muestra de jarosita.

## Dickita

El último espectro para ser analizado en esta sección de validación de la metodología, es el perteneciente a una muestra de dickita. Los resultados se presentan en la Figura 4.10, en la cuál, se puede observar que solamente los primeros dos espectros sugeridos corresponden a espectros de dickitas. Un análisis visual de los espectros en la Figura 4.10 sugiere que el espectro corregido corresponde a los dos espectros con un mayor coeficiente de correlación, debido a una mayor congruencia en las amplitudes de las frecuencias características.



**Figura 4.10:** Resultado de aplicar la metodología airPLS al espectro de una muestra de dickita.

## 4.5. Comparación entre las metodologías

De acuerdo con los resultados presentados anteriormente en este capítulo y la información de la Figura 4.11 haremos una descripción cualitativa entre los métodos explorados.

El método de regresión polinomial se destaca por obtener altos coeficientes de correlación (por encima del .80). A pesar de esto, se requieren tiempos de cómputo de entre 200 y 300 segundos para procesar cada muestra.

El método de filtro SG comparte similares tiempos de cómputo con la opción de regresión polinomial. A diferencia del método de regresión polinomial, este método no presenta resultados igual de consistentes. Incluso, para la muestra de Dickita obtuvo el coeficiente de correlación más bajo presentado en este capítulo.

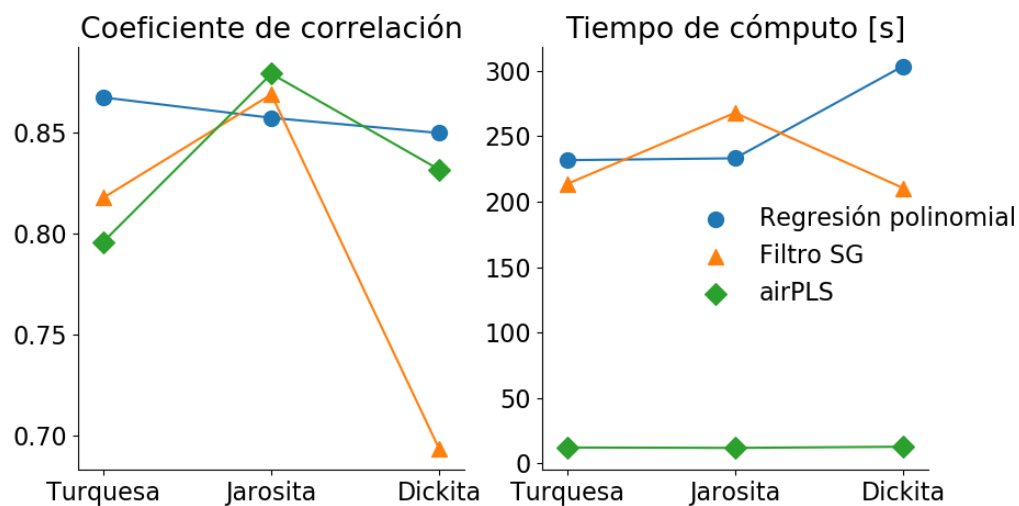
El método airPLS resultó ser el menos demandante en cuestión de tiempo de cómputo, cada muestra se procesa aproximadamente en 15 segundos. Además, los coeficientes de correlación obtenidos con airPLS son similares a los otros dos métodos.

Aunque el método de airPLS no resulte concluyente en un 100 % de las muestras, se recomienda utilizarse como primera opción, debido a que es el método que

#### 4. VALIDACIÓN DE LA METODOLOGÍA

---

requiere un menor tiempo de cómputo. En caso de no obtener un resultado decisivo, se recomienda utilizar el método polinomial o el filtro SG para comprobar los resultados u obtener una clasificación diferente.



**Figura 4.11:** Representación del tiempo de cómputo necesario y el coeficiente de correlación obtenido para las muestras analizadas mediante cada método.



---

## Capítulo 5

# Aplicación de la metodología:

# Proyecto del Instituto Nacional de Antropología e Historia

---

<b>5.1. Los purépecha</b> . . . . .	<b>50</b>
<b>5.2. Muestras arqueológicas</b> . . . . .	<b>51</b>
<b>5.3. Análisis espectral</b> . . . . .	<b>51</b>
5.3.1. Regresión polinomial . . . . .	51
5.3.2. Filtro Savitzky-Golay y señal envolvente . . . . .	55
5.3.3. Mínimos cuadrados penalizados con ponderación itera- tivamente adaptable (airPLS) . . . . .	58
<b>5.4. Comparación entre las metodologías</b> . . . . .	<b>60</b>

---

En este capítulo se aplican las tres metodologías examinadas en este trabajo (Regresión polinomial, Filtro Savitzky-Golay y Mínimos cuadrados penalizados con ponderación iterativamente adaptable) a muestras antropológicas que pertenecieron a la cultura de los purépechas. El uso de estas muestras es dentro del marco del proyecto del Instituto Nacional de Antropología e Historia para reconocer los minerales utilizados en los símbolos de poder purépecha. Además, se presenta el contexto histórico de los purépechas como civilización y su relación con los minerales asociados a las muestras analizadas.

## 5.1. Los purépecha

Para 1521, cuando los españoles llegaron a las regiones que dominaban los purépechas, la ciudad de Tzintzuntzan, colindando con el lago Pátzcuaro, era el núcleo político y religioso desde el que se controló una vasta región que abarcaba casi todo el estado que se conoce hoy como Michoacán, además abarcaba una parte significativa de los estados de Guanajuato y Guerrero; algunas otras pequeñas regiones controladas eran pertenecientes a los territorios de los estados de Jalisco, Querétaro y la Ciudad de México.

Sobre el origen de los purépechas, se sabe poco, los mexicas los llamaban michoaque (los de la tierra del pescado). Algunos piensan que es un grupo tardío ya que su estado hegemónico fue hasta el siglo XIII d.C. Existe una falta de evidencias que demuestren su migración reciente, como comunidades que compartan nombres que indiquen la ruta que, supuestamente, siguieron, o hallazgos arqueológicos que demuestren su camino en los sitios de la supuesta migración.

El imperio purépecha termina el 25 de julio de 1522 cuando Cristóbal de Olid, un capitán de Cortés, entra a Tzintzuntzan.

### El arte Purépecha

Como insignias de rango estaban los calabazos o guajes laqueados y engastados con turquesas que llevan los sacerdotes en la espalda. Piezas de oro y plata, en especial un tipo de pinzas colgadas en el pecho, lo mismo que los bezotes de obsidiana para ornamentar el mentón, que eran solo usados para las clases altas.

Al conocer tan poco sobre el origen de los purépecha y de los pueblos que los antecedieron, es muy difícil definir cuales fueron las artes que les correspondieron y cuáles las propias de sus tributarios. El panorama más que oscuro es problemático ya que el señorío purépecha fue un Estado multiétnico del cual se han realizado pocos trabajos arqueológicos hasta ahora.

La alfarería fue un arte-oficio destacado en el occidente de México; y los michoacanos, como pertenecientes a una parte de esta gran área, no podían quedar mal en esta rama. Casi todas las piezas en los museos proceden de entierros pertenecientes a la elite. De ahí quizá se deriva la calidad de su manufactura y acabado. (*Artes de México*, (2007))

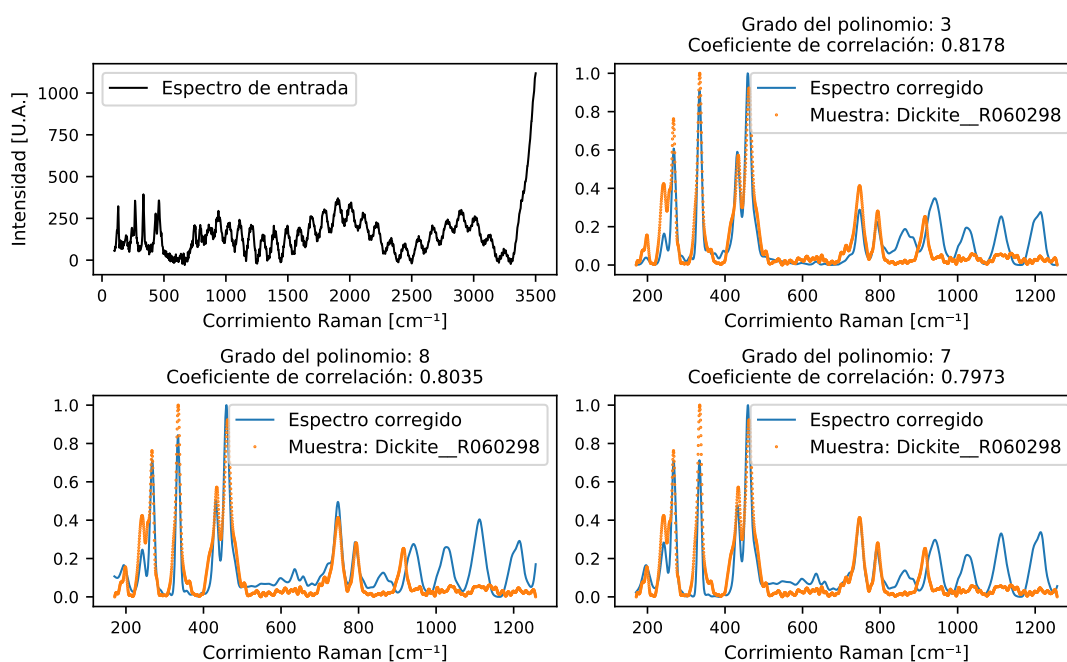
## 5.2. Muestras arqueológicas

Para ilustrar el potencial del trabajo realizado, se presentan las correcciones espectrales para tres de las muestras arqueológicas facilitadas por el Museo Regional Michoacano. Las piezas se relacionan con la cultura purépecha y sus símbolos de poder (ver Fig. 5.1).

## 5.3. Análisis espectral

Las muestras analizadas pertenecen a minerales de dickita, jadeíta y turquesa.

### 5.3.1. Regresión polinomial



**Figura 5.2:** Corrección del espectro del bezote de obsidiana con incrustación por el método de regresión polinomial.

En la Figura 5.2 se observa que las tres coincidencias que resultan del procesamiento por regresión polinomial corresponden a muestras de dickita. A pesar

5. APLICACIÓN DE LA METODOLOGÍA: PROYECTO DEL INSTITUTO NACIONAL DE ANTROPOLOGÍA E HISTORIA

---



(a) Bezote de obsidiana con incrustación lapidaria en forma de disco.



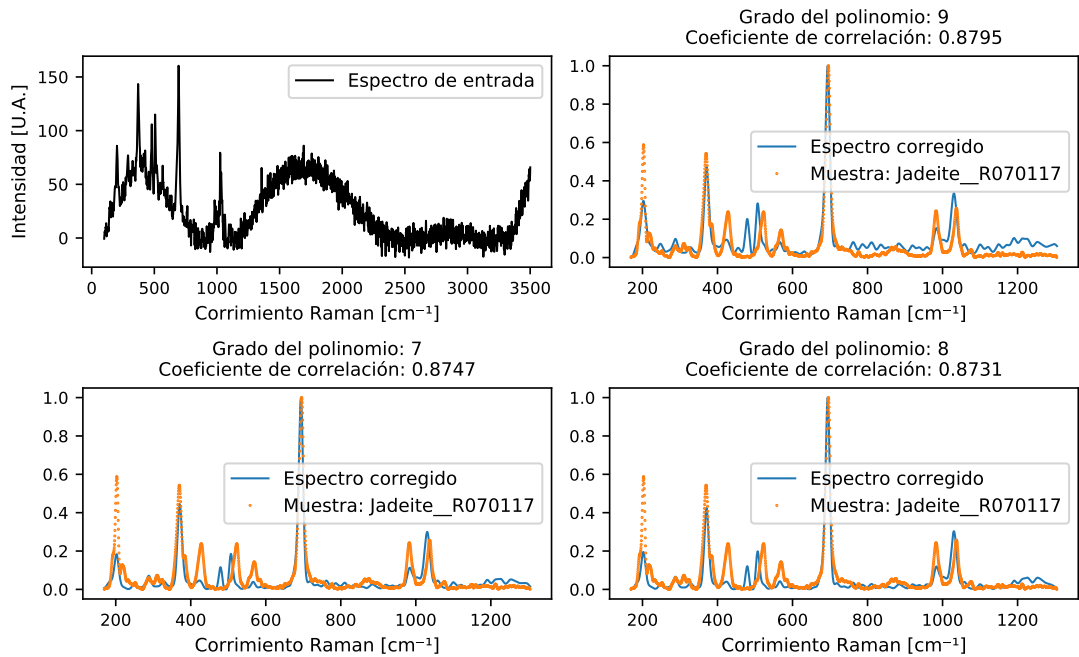
(b) Pendientes de jadeíta.



(c) Sartal de turquesas.

**Figura 5.1:** Símbolos de poder purépecha, principalmente utilizados por la nobleza

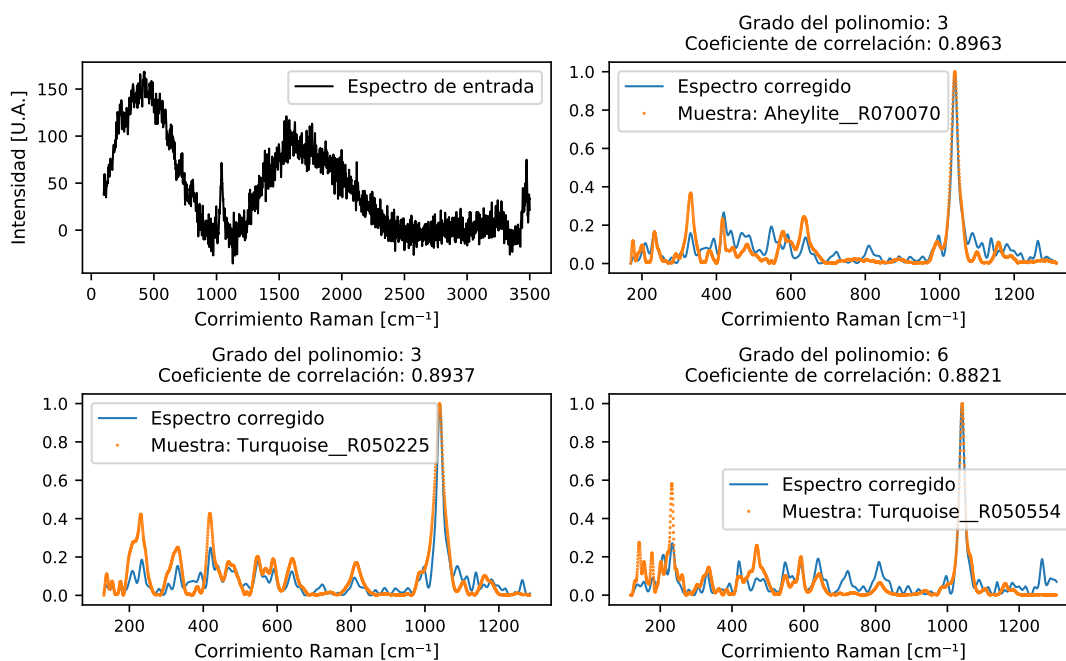
de que se obtienen tres coincidencias con altos coeficientes de correlación, es evidente que en altos números de onda ( $\sim 800 - 1200 \omega$ ) existe ruido que no pudo ser eliminado por la regresión polinomial.



**Figura 5.3:** Corrección del espectro de los pendientes de jadeíta por el método de regresión polinomial.

Para el espectro de los pendientes (ver Fig. 5.3) se obtienen tres coincidencias correspondientes con jadeíta. Si se recurre a un análisis visual se comprueba que las frecuencias características del espectro corregido coinciden con las de los espectros de referencia de la jadeíta.

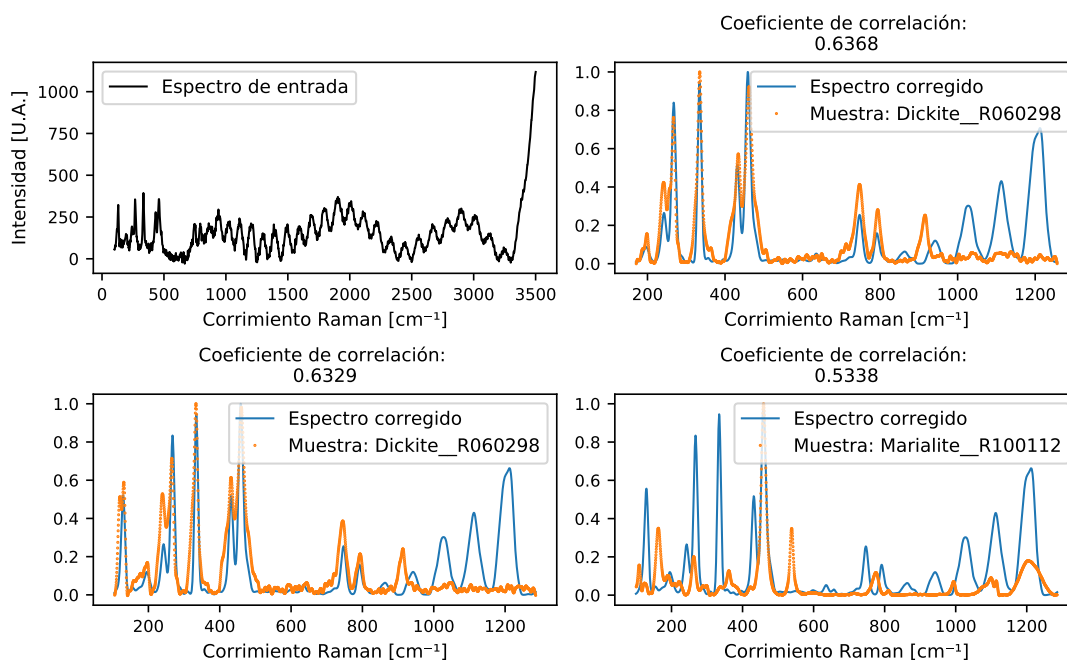
## 5. APLICACIÓN DE LA METODOLOGÍA: PROYECTO DEL INSTITUTO NACIONAL DE ANTROPOLOGÍA E HISTORIA



**Figura 5.4:** Corrección del espectro del sartal de turquesas por el método de regresión polinomial.

Para el análisis del sartal de turquesas, se trabajó con las cuentas sueltas que envuelven fragmentos de tibia y peroné (ver Fig. 5.1c). Como se puede observar de la Figura 5.4, únicamente dos de las tres coincidencias corresponden a muestras de turquesas, la similitud entre el espectro de la aheylita y la turquesa (ver Fig. 4.1), como se mencionó en la sección 4.2, propicia la aparición de la aheylita al analizar muestras de turquesa.

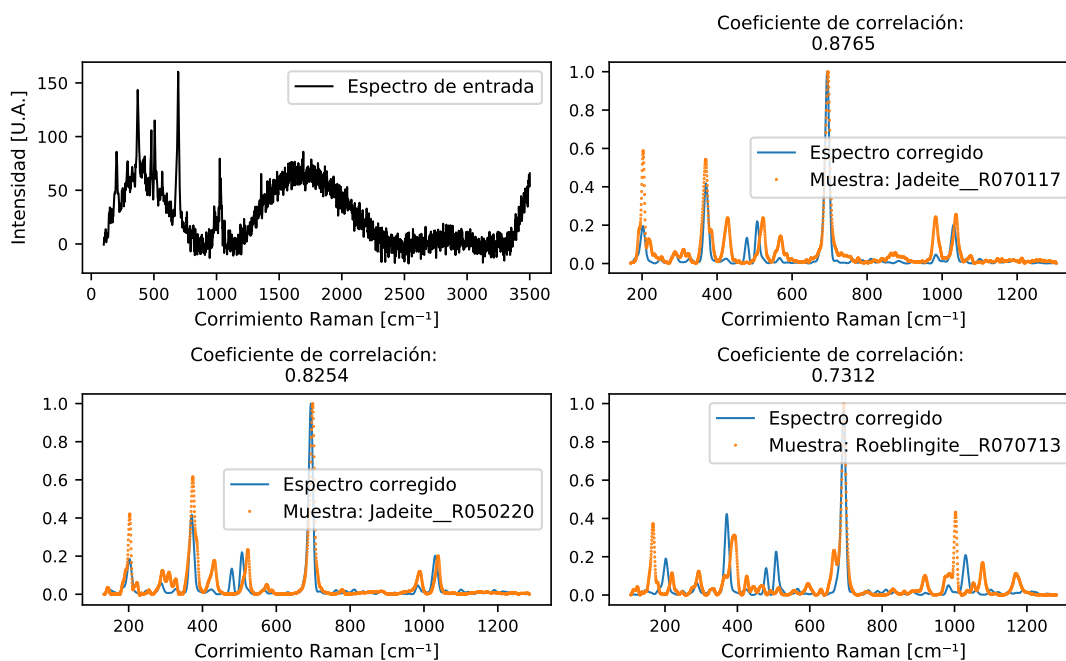
### 5.3.2. Filtro Savitzky-Golay y señal envolvente



**Figura 5.5:** Corrección del espectro del bezote de obsidiana con incrustación por el método de filtro Savitzky-Golay.

La muestra analizada del bezote de dickita, genera los resultados de la Figura 5.5, en la que se observa que las dos coincidencias con un mayor coeficiente de correlación corresponden a muestras de referencia de una dickita. La última coincidencia corresponde al espectro de una marialita, pero debe tomarse en cuenta que el coeficiente de correlación de este último espectro se encuentra muy por debajo de los primeros dos. Además, un análisis visual revela que solamente las dos primeras coincidencias tienen una buena correspondencia en frecuencias características con el espectro corregido.

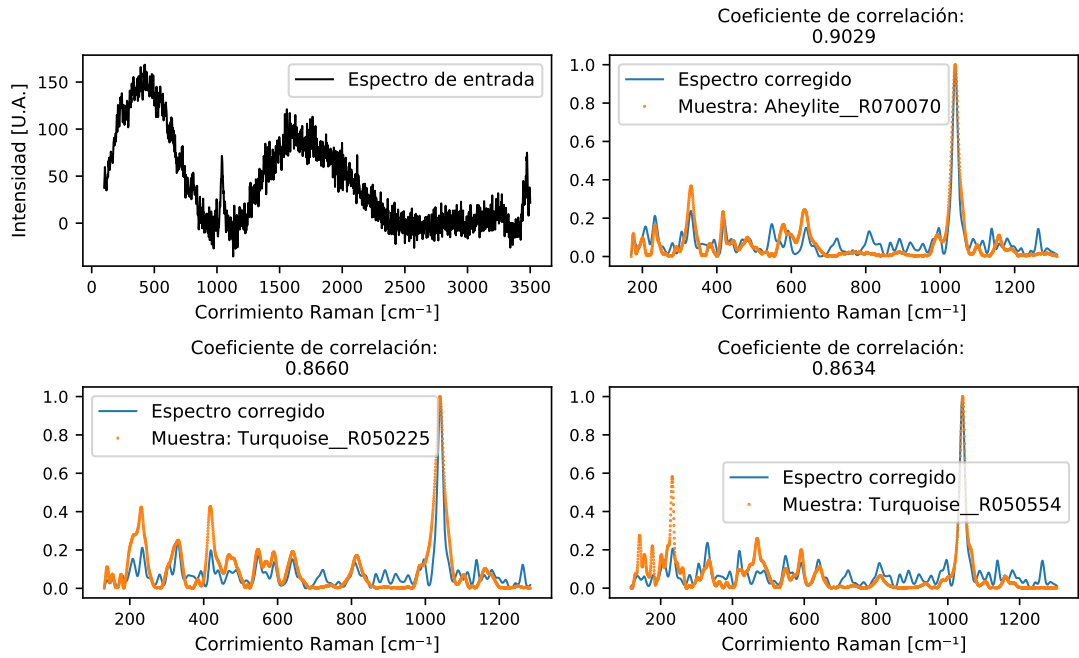
## 5. APLICACIÓN DE LA METODOLOGÍA: PROYECTO DEL INSTITUTO NACIONAL DE ANTROPOLOGÍA E HISTORIA



**Figura 5.6:** Corrección del espectro de los pendientes de jadeíta por el método de filtro Savitzky-Golay.

Para los pendientes de la Figura 5.1b se obtienen los resultados de la Figura 5.6, en la que se observan dos coincidencias de jadeíta y una de roeblingita. De nuevo, realizando un análisis visual se observa una mayor correlación en las frecuencias características de las jadeítas con el espectro corregido que la que presenta la roeblingita. Además, el coeficiente de correlación de la roeblingita se encuentra bastante por debajo de los coeficientes de las jadeítas.

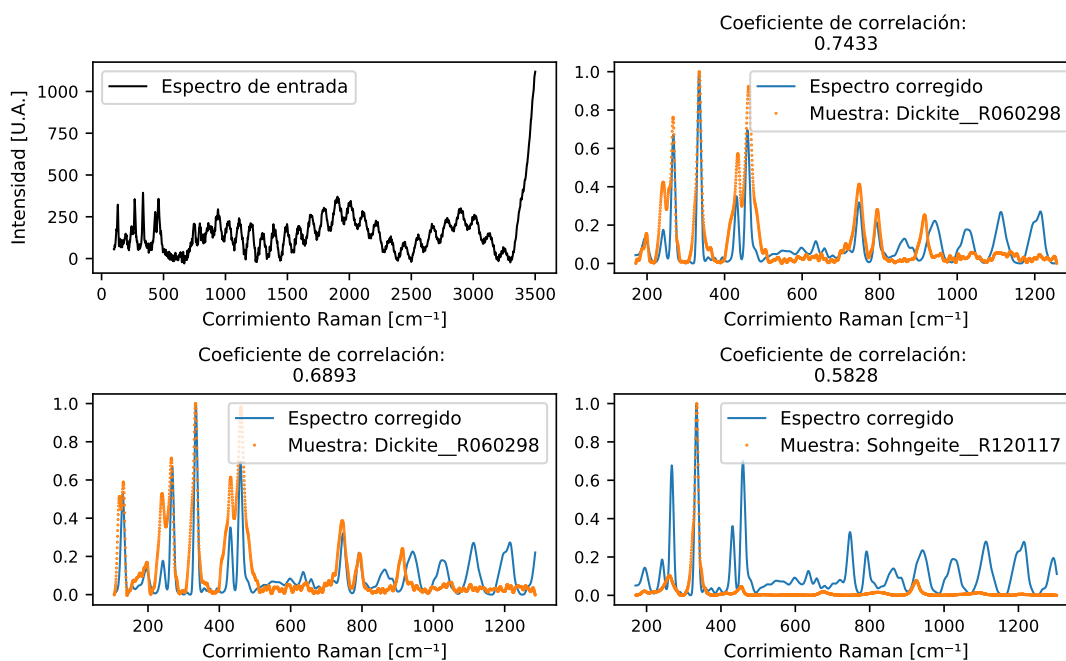




**Figura 5.7:** Corrección del espectro del sartal de turquesas por el método de filtro Savitzky-Golay.

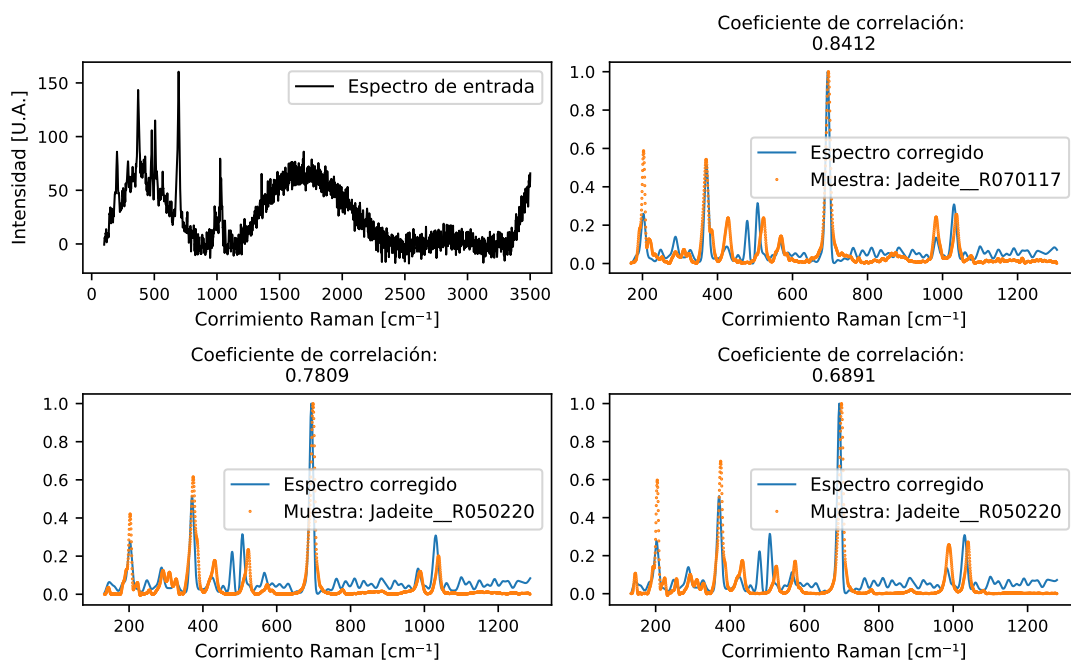
El análisis de las cuentas del sartal de la Figura 5.1c corresponde con los resultados de la Figura 5.7. Se pueden observar dos coincidencias para turquesas y una para aheylita, como se ha comentado, esto es resultado de la similitud entre los espectros de la aheylita y la turquesa.

### 5.3.3. Mínimos cuadrados penalizados con ponderación iterativamente adaptable (airPLS)



**Figura 5.8:** Corrección del espectro del bezote de obsidiana con incrustación por el método airPLS.

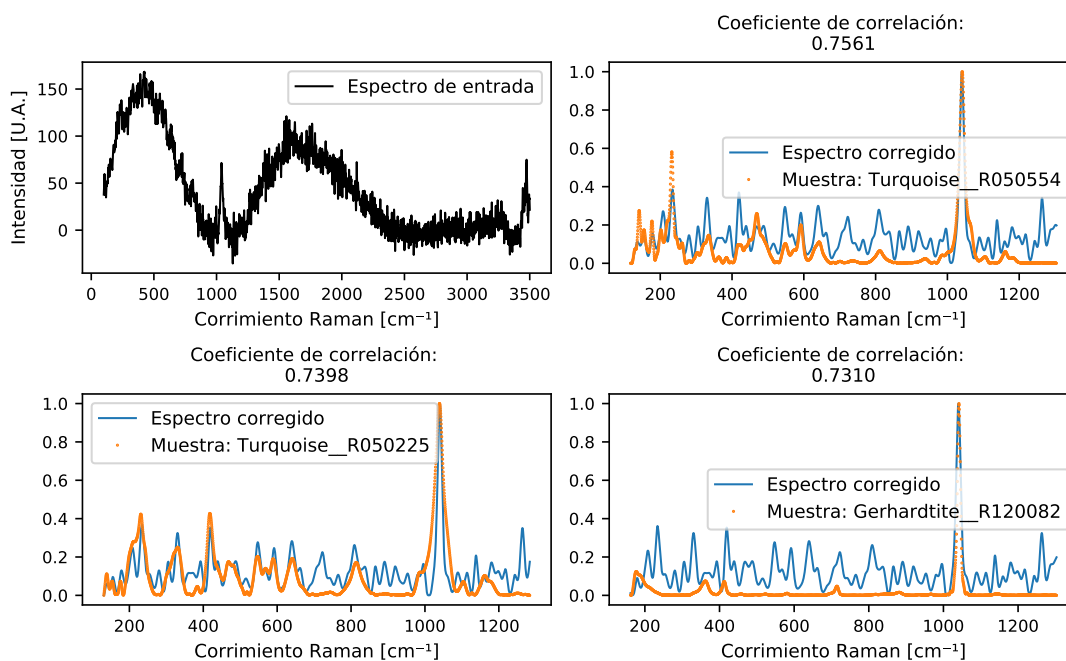
El bezote de dickita de la Figura 5.1a después de ser corregido con el método airPLS, arroja los resultados de la Figura 5.8. Se puede observar que los dos primeros espectros que coinciden con el espectro corregido, son los de muestras de dickitas. Por el contrario, la tercera coincidencia corresponde a una sohngeita, la cual se puede descartar con base en el análisis visual y al coeficiente de correlación tan bajo (0.5828) en comparación con los espectros de dickitas (0.6893 y 0.7433).



**Figura 5.9:** Corrección del espectro de los pendientes de jadeíta por el método airPLS.

El espectro corregido de los pendientes de la Figura 5.1b se muestra en la Figura 5.9, así como el espectro original de los pendientes y sus coincidencias con los espectros de referencia. En este caso, los tres espectros con un mayor coeficiente de correlación corresponden al mineral jadeíta.

## 5. APLICACIÓN DE LA METODOLOGÍA: PROYECTO DEL INSTITUTO NACIONAL DE ANTROPOLOGÍA E HISTORIA



**Figura 5.10:** Corrección del espectro del sartal de turquesas por el método airPLS.

Por último, en la Figura 5.10 se presentan los resultados para el espectro del sartal. En los dos primeros espectros de referencia, se observan dos coincidencias para turquesa y en el tercero una coincidencia para gerhardtita. El espectro de la gerhardtita se puede descartar debido a la escasa similitud que presenta con el espectro corregido. Si bien, los primeros dos espectros no son conclusivos, presentan una correlación mayor en las frecuencias características con el espectro corregido.

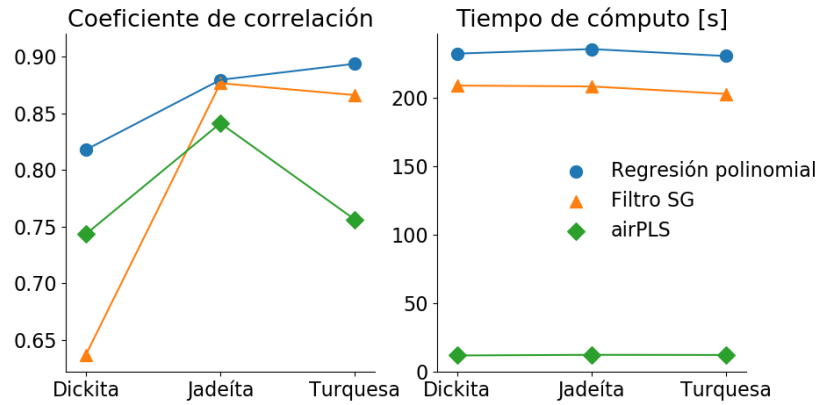
### 5.4. Comparación entre las metodologías

Según los resultados de la Figura 5.11, el método airPLS mantiene los menores tiempos de cómputo para corregir cada muestra.

El método de regresión polinomial encontró los mayores coeficientes de correlación en los tres casos analizados. A pesar de esto, el tiempo de cómputo necesario para corregir cada muestra con el método de regresión polinomial es de aproximadamente 220 segundos, seguido por los métodos de filtro SG (200 s) y airPLS (15 s).

En general, se recomienda utilizar el método que mejor ajuste las necesidades del estudio, dependiendo del volumen de datos a procesar. Si se dispone del

tiempo, se recomienda utilizar el método de regresión polinomial. En caso de que el volumen de datos no permita el análisis de cada muestra con el método de regresión polinomial, el método airPLS puede arrojar resultados preliminares aceptables.



**Figura 5.11:** Representación del tiempo de cómputo necesario y el coeficiente de correlación obtenido para las muestras analizadas mediante cada método.



---

## Capítulo 6

# Conclusiones y comentarios

---

El objetivo de esta tesis fue desarrollar herramientas que permitan determinar la clasificación de una muestra al procesar los espectros de Raman crudos (i.e. tal cual se obtienen del espectrómetro Raman). La principal dificultad era la eliminación de ruido en los espectros debido a la fluorescencia de las muestras. Con ese objetivo en mente, se enfocaron esfuerzos para reducir el efecto de la desviación de la línea base y del ruido en altas frecuencias. Se exploraron tres métodos para remover la desviación de la línea base y un único filtro digital para eliminar el ruido en altas frecuencias inherente a la medición. Los tres métodos explorados son: regresión polinomial, filtro Savitzky-Golay y mínimos cuadrados penalizados con ponderación iterativamente adaptable (airPLS). Utilizando los métodos propuestos, se consiguió recuperar, en general, espectros que arrojaron coincidencias adecuadas con los espectros de la base de datos RRUFF.

A partir de nuestros resultados, la corrección del efecto de la fotoluminiscencia en espectros ruidosos, mejora significativamente la probabilidad de encontrar una coincidencia con alguna base de datos de espectros ideales. La correcta eliminación del ruido espectral ayudará a los laboratoristas a catalogar y a describir de una manera precisa las muestras que sean del interés del IGUM.

De los tres métodos se recomienda el de regresión polinomial por ser más estable, si no se dispone del tiempo necesario o si el volumen a analizar es muy grande, se recomienda el uso del método airPLS por ser mucho más eficiente en cuestión de tiempo de cómputo. Por último, el método de filtro SG debe considerarse como un criterio adicional en caso de que ninguno de los dos métodos anteriores haya sido concluyente.

Un problema que no alcanza a explorar esta tesis es el de qué ocurre si la muestra analizada contiene más de una firma característica en el espectro de Raman. Se considera la posibilidad de continuar con este trabajo en el futuro desarrollando herramientas que permitan reconocer si el espectro analizado contiene más componentes moleculares, además del que predomina. Con el fin de conocer

## 6. CONCLUSIONES Y COMENTARIOS

---

completamente la composición de la muestra en espectros de esta naturaleza, es necesario aislar de su espectro cada componente que lo constituye.



---

Apéndice A

## Repositorio en GitHub

---

Con el fin de compartir los productos de esta Tesis, se habilita en GitHub el repositorio *Raman\_Processing*, donde se encuentra el procedimiento para reproducir los resultados, así como los códigos y datos necesarios.

[https://github.com/victoralexander132/Raman\\_Processing](https://github.com/victoralexander132/Raman_Processing)



## Referencias

---

- Bindi, L., Churakov, S. V., Downs, R., Gfeller, F., Krivovichev, S. V., Lafuente, B., Mugnaioli, E., Nestola, F., Steinhardt, P. J., Stone, N., et al. (2015). *Highlights in mineralogical crystallography*. Walter de Gruyter GmbH & Co KG. [13](#)
- Bocklitz, T., Walter, A., Hartmann, K., Rösch, P., y Popp, J. (2011). How to pre-process raman spectra for reliable and stable models? *Analytica chimica acta*, 704(1-2):47–56. [IX](#), [14](#)
- Cebeci-Maltaş, D. (2017). Photobleaching profile of raman peaks and fluorescence background. *Recurso en línea* <https://www.europeanpharmaceuticalreview.com/article/70503/raman-peaks-fluorescence-background/> [Consultado: 29 septiembre 2020]. [IX](#), [9](#)
- Ferraro, J. R. (2003). *Introductory raman spectroscopy*. Elsevier. [IX](#), [XIII](#), [5](#), [6](#)
- Fosbury, R. (2013). Fluorescence spectroscopy of minerals: 1—using a spectrometer. *Recurso en línea* <http://www.eso.org/~rfosbury/Article%20from%20Journal2013.pdf> [Consultado: 8 octubre 2020]. [IX](#), [8](#)
- Green, P. J. y Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall/CRC. [23](#)
- Konorov, S. O., Addison, C. J., Schulze, H. G., Turner, R. F., y Blades, M. W. (2006). Hollow-core photonic crystal fiber-optic probes for raman spectroscopy. *Optics letters*, 31(12):1911–1913. [2](#)
- Liang, Y.-Z., Leung, A. K.-M., y Chau, F.-T. (1999). A roughness penalty approach and its application to noisy hyphenated chromatographic two-way data. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 13(5):511–524. [23](#)

## REFERENCIAS

---

- McMillan, P. F. (1989). Raman spectroscopy in mineralogy and geochemistry. *Annual review of earth and planetary sciences*, 17(1):255–279. [1](#)
- Montgomery, D. C., Peck, E. A., y Vining, G. G. (2012). *Introduction to linear regression analysis*, volume 821. John Wiley & Sons. [14](#)
- Potts, P. J., Bowles, J., Reed, S. J., y Cave, R. (1995). *Microprobe techniques in the earth sciences*, volume 6. Springer Science & Business Media. [1](#)
- Raman, C. V. y Krishnan, K. S. (1928). A new type of secondary radiation. *Nature*, 121(3048):501. [1](#), [3](#)
- Rowlands, C. J. y Elliott, S. R. (2011). Denoising of spectra with no user input: a spline-smoothing algorithm. *Journal of Raman Spectroscopy*, 42(3):370–376. [2](#)
- Savitzky, A. y Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639. [20](#)
- Smekal, A. (1923). Zur quantentheorie der dispersion. *Naturwissenschaften*, 11(43):873–875. [3](#)
- Tarback, E. J., Lutgens, F. K., Tasa, D., y Linneman, S. (2005). *Earth: an introduction to physical geology*. Pearson/Prentice Hall Upper Saddle River. [ix](#), [6](#), [7](#)
- Artes de México, (2007). *Viaje por sendas purépechas*. Secretaría de Cultura de Michoacán. [50](#)
- Weisstein, E. W. (2002). Least squares fitting—polynomial. *Recurso en línea* <http://mathworld.wolfram.com/LeastSquaresFittingPolynomial.html> [Consultado: 30 septiembre 2019]. [15](#)
- Whittaker, E. T. (1922). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75. [23](#)
- Zhang, Z.-M., Chen, S., y Liang, Y.-Z. (2010a). Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135(5):1138–1146. [23](#), [44](#)
- Zhang, Z.-M., Chen, S., Liang, Y.-Z., Liu, Z.-X., Zhang, Q.-M., Ding, L.-X., Ye, F., y Zhou, H. (2010b). An intelligent background-correction algorithm for highly fluorescent samples in raman spectroscopy. *Journal of Raman Spectroscopy*, 41(6):659–669. [24](#), [25](#)