



**UNIVERSIDAD DE CONCEPCIÓN  
FACULTAD DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA**



**ANÁLISIS DE IMÁGENES HIPER ESPECTRALES EN EL RANGO INFRARROJO DE  
CONCENTRADOS DE COBRE USANDO HYPER-TOOLS**

**POR**

**Patricio Humberto Quezada Burgos**

Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción para  
optar al título profesional de Ingeniero Civil Electrónico

Profesor Guía  
Daniel Sbárbaro Hofer

Julio 2023  
Concepción (Chile)

© 2023 Patricio Humberto Quezada Burgos

©2023 Patricio Humberto Quezada Burgos

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

*A mis queridos padres, quienes han sido un pilar fundamental en mi desarrollo personal, brindándome su incondicional compañía y apoyando cada decisión que he tomado en mi vida.*

## Sumario

---

La minería es una industria que ha crecido a pasos agigantados en el transcurso de los años y que se visto beneficiada por los grandes avances tecnológicos que se han ido desarrollando hasta la actualidad. Estas nuevas tecnologías han promovido el interés de la comunidad científica he impulsado una amplia variedad de estudios sobre la extracción de recursos mineros en donde se busca para aumentar su eficiencia y niveles de productividad.

En este estudio se aplicaron técnicas de regresión que permitieron mediante la extracción y manipulación de datos de imágenes hiper-espectrales de concentrados de cobre, obtener modelos que permitan explicar la variabilidad de la concentración porcentual de la mineralogía de ciertas muestras. Para lograr este objetivo se utilizó un conjunto de datos correspondientes a 32 imágenes hiper-espectrales tomadas a muestras de concentrados de cobre. A las que se le realizo un exhaustivo análisis, con el que se logró tener una estructura de datos muchos más limpia.

El trabajo se realizó en el software de Matlab en conjunto con la herramienta Hypertools, las cuales permitieron procesar los diversos datos de interés a estudiar, en ellos se implementó la aplicación del método de regresión por mínimos cuadrados (PLS) y nos entregó como resultado los modelos para cada uno de los minerales que se analizó.

Los resultados obtenidos dan cuenta de la capacidad adaptativa del modelo de regresión PLS, el cual generó modelos que lograron estimar con una precisión aceptable ( $R^2 > 0.70$ ) la variabilidad de las concentraciones de minerales entre las que se destacó las respuestas obtenidas en los minerales: Bornita, Calcopirita, Enargita, Pirita y Óxido de hierro, los que presentaron además los mejores resultados en términos de precisión al momento de enfrentarse a datos fuera del conjunto de entrenamiento.

## Summary

---

Mining is an industry that has grown at an astonishing pace over the years and has been benefited by significant technological advancements that have been continuously developed up to the present day. These new technologies have fostered the interest of the scientific community and driven a wide array of studies on mining resource extraction, aiming to enhance its efficiency and productivity levels.

In this study, regression techniques were applied that allowed obtaining models through the extraction and manipulation of hyperspectral image data from copper concentrates, explaining the variability of the percentage concentration of mineralogy in certain samples. To achieve this goal, a dataset corresponding to 32 hyperspectral images taken from copper concentrate samples was used. An exhaustive analysis was conducted on them, resulting in a much cleaner data structure.

The work was carried out using Matlab software in conjunction with the Hypertools tool, enabling the processing of diverse data of interest for study. The implementation of the Partial Least Squares (PLS) regression method was applied, yielding models for each of the minerals analyzed.

The obtained results demonstrate the adaptive capacity of the PLS regression model, which generated models that accurately estimated the variability of mineral concentrations with an acceptable precision ( $R^2 > 0.70$ ). Notably, the responses obtained for the minerals Bornite, Chalcopyrite, Enargite, Pyrite, and Iron Oxide stood out, presenting the best results in terms of precision when confronted with data outside the training set.

## Abreviaciones

---

NIR	:	Infrarrojo Cercano, <i>Near-Infrared</i> .
IR	:	Infrarrojo, <i>InfraRed</i> .
PLS	:	Mínimos Cuadrados Parciales, <i>Partial Least Squares</i> .
PCA	:	Análisis de Componente Principal , <i>Principal Componente Analysis</i> .
PCR	:	Regresión por Componentes Principales , <i>Principal Component Regression</i> .
NLV	:	Numero de Variables Latentes , <i>Number of Latent Variables</i> .
HSI	:	Imagen Hiperspectral , <i>Hyperspectral Imaging</i> .
RGB	:	Rojo, Verde, Azul , <i>Red, Green, Blue</i>
CV	:	Validación Cruzada , <i>Cross Validation</i>

# Tabla de Contenidos

<b>Tabla de contenidos</b> . . . . .	<b>ix</b>
<b>1 Introducción</b> . . . . .	<b>1</b>
1.1 Introducción General . . . . .	1
1.2 Revisión Bibliográfica . . . . .	2
1.2.1 Conceptos básicos y Teoría de Imágenes Hiper-espectrales . . . . .	2
1.2.2 Procesamiento de imágenes hiper-espectrales . . . . .	3
1.2.3 Métodos de Regresión . . . . .	4
1.2.4 Aplicaciones del Uso de Imágenes Hiper-espectrales . . . . .	6
<b>2 Definición del Problema</b> . . . . .	<b>8</b>
2.1 Introducción . . . . .	8
2.2 Hipótesis de Trabajo . . . . .	8
2.3 Objetivos . . . . .	9
2.3.1 Objetivo General . . . . .	9
2.3.2 Objetivos Específicos . . . . .	9
2.4 Alcances y Limitaciones . . . . .	9
2.5 Metodología . . . . .	10
<b>3 Marco Teórico</b> . . . . .	<b>11</b>
3.1 Introducción . . . . .	11
3.2 Tipo de Imágenes . . . . .	11
3.2.1 Imagen de un Solo Canal . . . . .	12
3.2.2 Imágenes en Color . . . . .	12
3.2.3 Imágenes Multi-espectrales . . . . .	13
3.2.4 Imágenes Hiper-espectrales . . . . .	13
3.3 Tipos de Escaneo de Imágenes Hiper-espectrales . . . . .	14
3.4 Estructura de una Imagen Hiper-espectral . . . . .	15
3.5 Pre-procesamiento de Imágenes Hiper-espectrales . . . . .	16
3.5.1 Eliminación de Píxeles Muertos . . . . .	16
3.5.2 Eliminación de Spikes . . . . .	17

3.5.3	Eliminación de Fondo . . . . .	17
3.5.4	Pre-procesamiento Espectral . . . . .	18
3.6	Obtención de Información Cuantitativa de Imágenes Hiper-espectrales Mediante Modelos de Regresión . . . . .	18
3.7	Regresión por Mínimos Cuadrados Parciales (PLS) . . . . .	20
<b>4</b>	<b>Desarrollo . . . . .</b>	<b>21</b>
4.1	Introducción . . . . .	21
4.2	Datos . . . . .	21
4.3	Lectura de Datos . . . . .	22
4.4	Eliminación de Píxeles Muertos y Spikes . . . . .	23
4.5	Determinación del Área de Trabajo . . . . .	25
4.6	Construcción matriz <b>X</b> e <b>Y</b> . . . . .	26
4.7	Análisis de la Distribución Mineral de Concentrados de Cobre . . . . .	26
4.7.1	Bornita . . . . .	27
4.7.2	Covelina . . . . .	27
4.7.3	Calcopirita . . . . .	28
4.7.4	Enargita . . . . .	28
4.7.5	Molibdenita . . . . .	29
4.7.6	Pirita . . . . .	29
4.7.7	Cuarzo . . . . .	30
4.7.8	Óxido de Hierro . . . . .	30
4.7.9	Otros Metales . . . . .	31
4.8	Selección de Parámetro para el Modelo PLS . . . . .	31
<b>5</b>	<b>Resultados . . . . .</b>	<b>32</b>
5.1	Modelo PLS para Bornita ( $\text{Cu}_5\text{FeS}_4$ ) . . . . .	32
5.2	Modelo PLS para Covelina ( $\text{CuS}_2$ ) . . . . .	33
5.3	Modelo PLS para Calcopirita ( $\text{CuFeS}_2$ ) . . . . .	34
5.4	Modelo PLS para Enargita ( $\text{Cu}_3\text{AsS}_4$ ) . . . . .	35
5.5	Modelo PLS para Molibdenita ( $\text{MoS}_2$ ) . . . . .	36
5.6	Modelo PLS para Pirita ( $\text{FeS}_2$ ) . . . . .	37



5.7	Modelo PLS para Cuarzo ( $\text{SiO}_2$ ) . . . . .	38
5.8	Modelo PLS para Oxido de Hierro ( $\text{Fe}_x\text{O}_y$ ) . . . . .	39
5.9	Modelo PLS para Otros Metales . . . . .	40
<b>6</b>	<b>Conclusiones . . . . .</b>	<b>41</b>
6.1	Discusión . . . . .	41
6.2	Conclusión . . . . .	41
6.3	Trabajo Futuro . . . . .	42

# Lista de Tablas

- Tabla 5.1.1: Rendimiento Modelo PLS Bornita . . . . . 32
- Tabla 5.2.1: Rendimiento Modelo PLS Covelina . . . . . 33
- Tabla 5.3.1: Rendimiento Modelo PLS Calcopirita . . . . . 34
- Tabla 5.4.1: Rendimiento Modelo PLS Enargita . . . . . 35
- Tabla 5.5.1: Rendimiento Modelo PLS Molibdenita . . . . . 36
- Tabla 5.6.1: Rendimiento Modelo PLS Pirita . . . . . 37
- Tabla 5.7.1: Rendimiento Modelo PLS Cuarzo . . . . . 38
- Tabla 5.8.1: Rendimiento Modelo PLS Oxido de Hierro . . . . . 39
- Tabla 5.9.1: Rendimiento Modelo PLS Otros Metales . . . . . 40

# Lista de Figuras

Figura 3.2.1: Imagen representativa de un píxel en una imagen. . . . .	11
Figura 3.2.1.1 : Imagen de una muestra de yogurt obtenida mediante Microscopio Confocal Láser de Barrido (CLSM) [19]. . . . .	12
Figura 3.2.2.1 : Imagen en color real (RGB) de una mariposa y su representación en mono-canales rojo verde y azul [20]. . . . .	12
Figura 3.2.3.1 : Imagen multi-espectral tomada de un billete de 10 euros [18]. . . . .	13
Figura 3.2.4.1 : Representación de la imagen de una galleta medida con una cámara hiper-espectral.[18]. . . . .	13
Figura 3.3.1 : Representación esquemática de las tres configuraciones comunes . . . . .	14
Figura 3.4.1 : Representación esquemática de la estructura del hipercubo en los tres métodos . . . . .	15
Figura 3.5.1.1 : Extracción de píxeles muertos de una muestra [25] . . . . .	16
Figura 3.5.2.1 : Ejemplo de eliminación de spikes [30] . . . . .	17
Figura 3.5.3.1 : 3 métodos de eliminación de fondo [30] . . . . .	17
Figura 3.5.4.1 : Métodos pre-procesamiento espectral [25] . . . . .	18
Figura 3.6.1 : Aplicación del modelado de regresión a imágenes hiper espectrales [22] .	19
Figura 3.6.2 : Aplicación del modelo de regresión al hipercubo, resultando en una imagen de predicción [22]. . . . .	19
Figura 4.3.1 : Ejemplo de visualización de imagen (Muestra B002). . . . .	22
Figura 4.4.1 : 20 espectros más representativos de muestra B002 . . . . .	23
Figura 4.4.2 : 20 Espectros mas representativos de Muestra B002 pre-procesada . . . . .	24
Figura 4.4.3 : Muestra B002 pre-procesada . . . . .	24
Figura 4.5.1 : Determinación del área de trabajo (Muestra B002) . . . . .	25
Figura 4.5.2 : Espectros mas representativos (Muestra B002), luego se seleccionar área de trabajo . . . . .	26
Figura 4.7.1.1 : Histograma de Bornita . . . . .	27
Figura 4.7.2.1 : Histograma de Covelina . . . . .	27
Figura 4.7.3.1 : Histograma de Calcopirita . . . . .	28
Figura 4.7.4.1 : Histograma de Enargita . . . . .	28
Figura 4.7.5.1 : Histograma de Molibdenita . . . . .	29

Figura 4.7.6.1 : Histograma de Pirita . . . . .	29
Figura 4.7.7.1 : Histograma de Cuarzo . . . . .	30
Figura 4.7.8.1 : Histograma de Óxido de Hierro . . . . .	30
Figura 4.7.9.1 : Histograma de Otros Metales . . . . .	31
Figura 5.1.1: Modelo PLS Bornita . . . . .	32
Figura 5.2.1: Modelo PLS Covelina . . . . .	33
Figura 5.3.1: Modelo PLS Calcopirita . . . . .	34
Figura 5.4.1: Modelo PLS Enargita . . . . .	35
Figura 5.5.1: Modelo PLS Molibdenita . . . . .	36
Figura 5.6.1: Modelo PLS Pirita . . . . .	37
Figura 5.7.1: Modelo PLS Cuarzo . . . . .	38
Figura 5.8.1: Modelo PLS Oxido de Hierro . . . . .	39
Figura 5.9.1: Modelo PLS Otros Metales . . . . .	40

# **1. Introducción**

---

## **1.1. Introducción General**

La minería ha jugado un rol importante en el progreso de la humanidad, desde tiempos remotos las diferentes comunidades que habitan el planeta han dedicado su tiempo a la búsqueda y extracción de minerales, los que le han permitido progresar y lograr el desarrollo de la sociedad como la conocemos hoy en día.

En la actualidad la actividad minera continúa teniendo un gran impacto a nivel mundial, es por esto que la industria siempre se encuentra en búsqueda de nuevos avances o desarrollo de nuevas tecnologías que le permitan mejorar u optimizar su proceso de extracción. Es en esta línea que surge la espectroscopia y uso de las imágenes hiper-espectrales como una potencial herramienta a desarrollar.

El estudio de las imágenes hiper-espectrales se ha desarrollado por muchos años y cada vez aumenta más el interés por su conocimiento, por lo que ha logrado abarcar diversas ramas de aplicación, debido a la gran cantidad de información que se encuentra contenida en ella. Considerando esto es preciso señalar que el desarrollo del presente trabajo se centrará principalmente en el ámbito de la minería, poniendo su énfasis en el análisis de concentrados de cobre mediante la toma de imágenes en el espectro.

En este trabajo se propone formular y aplicar un modelo de regresión matemático que permita la estimación de la mineralogía de concentrados de cobre a partir de la toma de imágenes hiper-espectrales, cuya manipulación de datos nos permita estimar la composición mineralógica de las muestras a analizar.

## 1.2. Revisión Bibliográfica

La información que se expondrá en esta sección corresponde a una revisión bibliográfica, obtenida a partir de diversos documentos como papers, libros, tesis de grado, revistas científicas, etc. Todo el material consultado se encuentra relacionado con imágenes Hiper-espectrales, su procesamiento y técnicas aplicadas para extraer su información, con el fin de llevar a cabo los objetivos del proyecto.

### 1.2.1. Conceptos básicos y Teoría de Imágenes Hiper-espectrales

- G. ElMasry y D.-W. Sun, “CHAPTER 1 - Principles of Hyperspectral Imaging Technology”, en *Hyperspectral Imaging for Food Quality Analysis and Control*, D.-W. Sun, ed., San Diego: Academic Press, 2010, págs. 3-43, ISBN: 978-0-12-374753-2. DOI: <https://doi.org/10.1016/B978-0-12-374753-2.10001-2>. dirección: <https://www.sciencedirect.com/science/article/pii/B9780123747532100012> [1]

Este capítulo del libro proporciona una visión general de las imágenes hiper-espectrales. En él se presentan los conceptos básicos, terminología y características que poseen este tipo de imágenes, así como la cantidad y calidad de la información que puede proporcionar y la instrumentación necesaria para su adquisición. Además, profundiza respecto a cuáles serían las principales ventajas y desventajas del trabajo con este tipo de datos, y se enfatiza en el potencial que puede llegar a obtener esta técnica.

- J. Qin, M. Kim, K. Chao et al., “Line-Scan Hyperspectral Imaging Techniques for Food Safety and Quality Applications”, *Applied Sciences*, vol. 7, pág. 125, ene. de 2017. DOI: [10.3390/app7020125](https://doi.org/10.3390/app7020125) [2]

Este artículo revisa las técnicas utilizadas para la obtención de imágenes hiper espectrales mediante escaneo lineal. Se abarcan tanto las técnicas existentes como también las emergentes, con el objetivo de proporcionar una visión completa de las posibilidades que ofrecen estas tecnologías.

Los temas principales incluyen técnicas de espectroscopia, métodos de calibración del sistema y técnicas de análisis espectrales, cada una desglosada y explicada en detalle para comprender su funcionamiento y aplicaciones específicas.

- A. Roman-Gonzalez y N. I. Vargas-Cuentas, “Análisis de imágenes hiperespectrales”, *Revista Ingeniería & Desarrollo*, vol. 9, n.º 35, págs. 14-17, 2013 [3]

Este artículo proporciona una visión general en cuanto a la temática sobre imágenes hiperespectrales, abordando aspectos teóricos, la composición de estas imágenes y además consideraciones importantes que se deben tener en cuenta al utilizar esta tecnología.

En el documento también se estudia cómo afecta la composición de los elementos ante la presencia de radiación y se ofrece una visión superficial de las aplicaciones y clasificaciones de las imágenes hiper-espectrales, según las necesidades específicas de cada aplicación. Esta información permite tener un primer acercamiento y adquirir los conocimientos básicos necesarios para posteriormente profundizar más en el tema.

### 1.2.2. *Procesamiento de imágenes hiper-espectrales*

- J. Amigo, H. Babamoradi y S. Elcoroaristizabal, “Hyperspectral image analysis. A tutorial”, *Analytica Chimica Acta*, vol. 896, págs. 34-51, 2015. DOI: 10.1016/j.aca.2015.09.030. dirección: <https://doi.org/10.1016/j.aca.2015.09.030> [4]

Artículo elaborado a modo de tutorial, que tiene como objetivo abordar temas como la adquisición y estructura de imágenes hiper-espectrales, el procesamiento de estas imágenes, el análisis exploratorio multivariado, la clasificación y el procesamiento final que debe tener, exponiendo pautas y herramientas prácticas que ayudaran en el análisis.

- Å. Rinnan, F. van den Berg y S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra”, *TrAC Trends in Analytical Chemistry*, vol. 28, n.º 10, págs. 1201-1222, 2009, ISSN: 0165-9936. DOI: <https://doi.org/10.1016/j.trac.2009.07.007>. dirección: <https://www.sciencedirect.com/science/article/pii/S0165993609001629> [5]

Esta revisión describe y compara los fundamentos teóricos y algorítmicos de los métodos de procesamiento actuales, los que se señalan se pueden dividir en dos categorías: métodos de corrección de dispersión y derivados espectrales. En él se señalan las consecuencias cualitativas y cuantitativas de su aplicación, con el objetivo de proporcionar a los usuarios NIR mejores modelos finales a través de un conocimiento fundamental acerca del preprocesamiento espectral.

- J.-C. Boulet y J.-M. Roger, “Pretreatments by means of orthogonal projections”, *Chemometrics and Intelligent Laboratory Systems*, vol. 117, págs. 61-69, 2012, Special Issue Section: Selected Papers from the 1st African-European Conference on Chemometrics, Rabat, Morocco, September 2010 Special Issue Section: Preprocessing methods Special Issue Section: Spectroscopic imaging, ISSN: 0169-7439. DOI: <https://doi.org/10.1016/j.chemolab.2012.02.002>. dirección: <https://www.sciencedirect.com/science/article/pii/S0169743912000251> [6]

En este artículo se describen varios pre-tratamientos lineales basados en proyecciones ortogonales. Enfocando su contenido en la forma que se identifica la información no deseada de una imagen hiper-espectral, utilizando conjuntos de datos de calibración, espectros puros, diseños experimentales o modelos matemáticos. el objetivo de eliminar toda la información innecesaria es tener una imagen lo más limpia posible. Así la corrección debe realizarse con una única proyección euclidiana.

- Å. Rinna, “Pre-processing in vibrational spectroscopy – when, why and how”, *Analytical Methods*, vol. 6, pág. 7124, 2014. DOI: [10.1039/C3AY42270D](https://doi.org/10.1039/C3AY42270D) [7]

Este documento se señala la importancia en procesar los datos de manera correcta puesto que si esto no es realizado produce confusión entre la información que se busca y el ruido que no es de interés. Este último factor es uno de los aspectos que toma mayor relevancia al tratar con imágenes hiper-espectrales, puesto que no está compuesto solo de desviaciones aleatorias de la propia medición, si no que está formado por variaciones como la dispersión de la luz.

Así el enfoque se centrará en prácticas de procesamiento y en aspectos que se debe tener en cuenta al utilizar técnicas NIR, IR y espectroscopia Raman.

### 1.2.3. Métodos de Regresión

- C. E. Alciaturi, M. E. Escobar, C. De La Cruz et al., “Partial least squares (PLS) regression and its application to coal analysis”, es, *Revista tecnica de la Facultad de Ingenieria Universidad del Zulia*, vol. 26, págs. 197-204, dic. de 2003, ISSN: 0254-0770. dirección: [http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S0254-077020030003000006&nrm=iso](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0254-077020030003000006&nrm=iso) [8]



Artículo centrado principalmente en el método de mínimos cuadrados parciales o partial least squares (PLS), en el que se destacan las ventajas teóricas y computacionales respecto a otros métodos, que lo han llevado a posicionarse como uno de los más utilizados en la quimiometría.

El artículo realiza una exhaustiva explicación del algoritmo matemático en el cual se basa el modelo para finalmente realizar un ejemplo de aplicación y verificar la efectividad de este.

- L. R., A. . y A. ., “Estudio comparativo de modelos de predicción obtenidos para propiedades de calidad en suelo, frutas y vegetales utilizando las técnicas de análisis multivariantes PLS y PCR”, Español, *Revista Ciencias Técnicas Agropecuarias*, vol. 18, págs. 50-54, 2009, ISSN: 1010-2760. dirección: <https://www.redalyc.org/articulo.oa?id=93215240011> [9]

Investigación se enfoca en comparar dos técnicas multivariantes, por un lado, el método de mínimos cuadrados parciales (PLS) y por otro el método de regresión de componentes principales (PCR). Para esto, ambos métodos multivariantes se aplicaron para medir la calidad del suelo, fruta y vegetales. Obteniendo como resultado que ambos tienen el potencial para estimar con precisión la calidad de los elementos evaluados y se concluye que el modelo PLS logra resultados ligeramente superiores.

- J. C. Vega-Vilca y J. Guzmán, “Regresión PLS y PCA como solución al problema de multicolinealidad en regresión múltiple”, *Rev. Mat.*, vol. 18, n.º 1, págs. 09-20, 2011, Electronic Document Format(ISO), ISSN: 1409-2433. dirección: [http://www.scielo.sa.cr/scielo.php?script=sci\\_arttext&pid=S1409-24332011000100002&lng=en&nrm=iso](http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1409-24332011000100002&lng=en&nrm=iso) [10]

En el artículo se presentan y comparan las técnicas análisis componentes principales (PCA) y la regresión por mínimos cuadrados parciales (PLS), como solución al problema acerca de la multi-colinealidad al momento de realizar regresiones múltiples. En el documento se explica detalladamente la teoría y fundamentos matemáticos detrás de ambos modelos, para finalmente realizar una aplicación práctica en donde se concluye la superioridad de la técnica PLS.

- P. D. Wentzell y L. Vega Montoto, “Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures”, *Chemometrics and Intelligent Laboratory Systems*, vol. 65, n.º 2, págs. 257-279, 2003, ISSN: 0169-7439. DOI: [https://doi.org/10.1016/S0169-7439\(02\)00138-7](https://doi.org/10.1016/S0169-7439(02)00138-7). dirección: <https://www.sciencedirect.com/science/article/pii/S0169743902001387> [11]

En este artículo se comparan dos de los métodos de calibración: la regresión por componentes principales (PCR) y la regresión por mínimos cuadrados parciales (PLS). En el documento se exponen los resultados de ambos métodos, mostrando como varían en errores de predicción y el número de variables latentes (NLV). En general en el estudio no se encontraron diferencias significativas en los errores de predicción de los métodos sin embargo se apreció que en la mayoría de los casos PLS requirió menos variables latentes.

#### 1.2.4. Aplicaciones del Uso de Imágenes Hiper-espectrales

- Á. F. Egaña, F. A. Santibáñez-Leal, C. Vidal et al., “A Robust Stochastic Approach to Mineral Hyperspectral Analysis for Geometallurgy”, *Minerals*, vol. 10, n.º 12, 2020, ISSN: 2075-163X. DOI: 10.3390/min10121139. dirección: <https://www.mdpi.com/2075-163X/10/12/1139> [12]

En este trabajo se presenta una nueva metodología para la caracterización de muestras metalúrgica. Este nuevo método surge ya que metodología clásica para el procesamiento de las imágenes hiper-espectrales utilizado en la industria no considera la variabilidad inherente de las características espectrales debido a las condiciones ambientales o la zona geográfica en donde la planta realiza su producción. Es por esto que en trabajo se plantea esta nueva metodología compuesta básicamente por: (a) un nuevo algoritmo de segmentación de imágenes hiper-espectrales, (b) un esquema de reducción de dimensionalidad que preserva información y (c) un modelo de regresión jerárquico estocástico. Finalmente se realiza una serie de experimento y se concluye que este modelo es capaz de adaptarse mejor a las variaciones del mundo real y producir predicciones mas confiables.

- C. Schropp, H. Kanpp y K. K. H. Neubert, “Potential of NIR hyperspectral imaging in the minerals industry”, en *OCM 2013 - Optical Characterization of Materials - conference proceedings / edited by Jürgen Beyerer / Fernando Puente León / Thomas Längle*, Karlsruhe: KIT Scientific Publishing, 2013, págs. 215-225. dirección: <https://publications.rwth-aachen.de/record/208866> [13]

En el presente documento se describe una investigación realizada en muestras de mineral de cobre proveniente de la mina kajaran, Armenia. Durante el estudio se utilizan dos técnicas de análisis para la caracterización, Análisis de liberación mineral y NIR HSI utilizando sistemas de imágenes hiper-espectrales SisUCHEMA (SPECIM) y el sistema de imágenes FTIR Spotlight 400 de Perkin

Elmer. Los resultados se comparan con técnicas de medición convencionales y se concluye que NIR HSI se puede convertir en una buena alternativa a los métodos analíticos convenciones, debido a sus múltiples formas de aplicación.

- C. Sandoval-Muñoz, G. Velásquez, J. Álvarez et al., “Enhanced elemental and mineralogical imaging of Cu-mineralized rocks by coupling  $\mu$ -LIBS and HSI”, *J. Anal. At. Spectrom.*, vol. 37, págs. 1981-1993, 10 2022. DOI: 10.1039/D2JA00147K. dirección: <http://dx.doi.org/10.1039/D2JA00147K> [14]

En este escrito se aborda una caracterización mineralógica detallada mediante la combinación de imágenes basadas en espectroscopia de emisión inducida por láser a micro-escala e imágenes hiper-espectrales en tres rocas mineralizadas con cobre diferentes. En él se caracterizaron la distribución de especies valiosas como la calcopirita, bornita, calcosina, covelina, enargita, magnetita y minerales estériles. Los resultados muestran que la caracterización elemental y mineralogía de las rocas mineralizadas con cobre puede llevarse a cabo mediante la combinación de ambas técnicas y permitiendo avanzar en técnicas que permita analizar rocas portadoras de cobre de manera automatizadas.

## **2. Definición del Problema**

---

### **2.1. Introducción**

El surgimiento constante de avances científicos y la creación de nuevas tecnologías es un ámbito que impacta significativamente en la industria minera. La implementación e incorporación de estos nuevos recursos, tanto en el proceso de extracción como en el procesamiento de estos materiales, es fundamental para que el desarrollo de la industria pueda progresar como lo ha venido haciendo hasta la actualidad. Para continuar en esa línea y aprovechar los grandes avances presentados en materia de espectrometría, se propone desarrollar un método de análisis de concentrado basado en el uso de imágenes hiper-espectrales y la aplicación del algoritmo Partial Least Squares (PLS), con el fin de estimar la composición mineralógica del contenido.

### **2.2. Hipótesis de Trabajo**

Como hipótesis general de este proyecto, planteamos se requiere de un método que permita el procesamiento de datos obtenidos a partir de imágenes Hiper-espectrales, con el que se pueda obtener una estimación de la mineralogía de concentrados de cobre.

Además, podemos plantear hipótesis específicas:

- Las imágenes Hiper-espectrales, tienen la capacidad de almacenar grandes cantidades de información por lo que un manejo adecuado de estas, beneficiaría de manera directa la eficiencia de los procesos productivos en la minería.
- La industria minera se encuentra en constante cambio, por lo que el aprovechamiento de esta nueva tecnología es de utilidad para su continuo crecimiento.
- En relación con los avances tecnológicos tenemos que la capacidad de procesamiento de datos de las maquinas ha aumentado con creces, por lo que es importante aprovechar estas ventajas y generar algoritmos más robustos y eficientes.

## **2.3. Objetivos**

### *2.3.1. Objetivo General*

A partir de imágenes Hiper-espectrales y mediante la utilización de software de ingeniería Matlab en conjunto con Hypertools y el uso del un algoritmo de regresión PLS, para su análisis, lograr obtener modelos que nos permita capturar la variabilidad de las concentraciones de lo diferentes minerales contenidos en el set de datos a analizar

### *2.3.2. Objetivos Específicos*

- Comprender y analizar la base de datos proporcionada.
- Realizar un correcto uso de la herramienta Hypertools.
- Realizar un pre-procesamiento de calidad de los datos proporcionados.
- Seleccionar de manera adecuada parámetros a utilizar para el tratamiento final de los datos.

## **2.4. Alcances y Limitaciones**

Dentro de las limitaciones de esta investigación, es importante señalar que se dependerá exclusivamente de datos de imágenes hiper-espectrales previamente capturados. Esto implica que no se requerirá el uso de instrumentos espectroscópicos durante el desarrollo del estudio. No obstante, esta dependencia de datos preexistentes también conlleva ciertas implicaciones en cuanto a la calidad y pertinencia de los datos capturados.

Las imágenes hiper espectrales que se ocuparan en el análisis están contenidas en archivos de formato de extensión ".bin", por lo que será necesario procesarlos de manera tal que nos permita trabajar con ellos tanto en el software de MATLAB como en la herramienta Hypertools.

El objetivo es aprovechar de manera óptima el gran potencial que ofrece el trabajo con imágenes Hiper-espectrales, para generar modelos de estimación de la concentración mineralógica en concentrados de cobre. Posteriormente, se evaluarán y analizarán los resultados obtenidos a partir de los modelos creados.

## **2.5. Metodología**

El desarrollo de este trabajo se está realizando con la ayuda del software MATLAB [15] y la herramienta Hypertools [16]. Ambas se ocuparán de manera paralela según sea el requerimiento necesario, puesto que la herramienta Hypertools corresponde a una GUI elaborada directamente en MATLAB, esta utiliza el mismo lenguaje y librerías existentes dentro del entorno del software, por lo que tendrá compatibilidad.

El conjunto de datos a utilizar como se mencionó fue obtenido desde una investigación previa[17] y en él se centrará todo el análisis y aplicación de técnicas de pre- y pos procesamiento como también técnicas de regresión, que permitan obtener una respuesta lo mas eficiente que se permita.

### 3. Marco Teórico

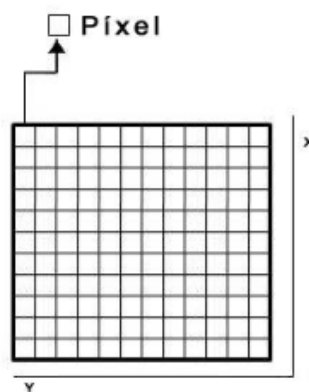
---

#### 3.1. Introducción

Este capítulo estará centrado primeramente en entender la procedencia y naturaleza de de los datos con los que se trabajará, además de proporcionar la teoría y los conceptos asociados a las imágenes Hiper-espectrales, dentro de los que se incluye su proceso de adquisición, características y estructura general. A través de esta exposición de contenido, se busca proporcionar una comprensión sólida de la base teórica necesaria para abordar adecuadamente el análisis y procesamiento de imágenes Hiper-espectrales en el contexto del proyecto.

#### 3.2. Tipo de Imágenes

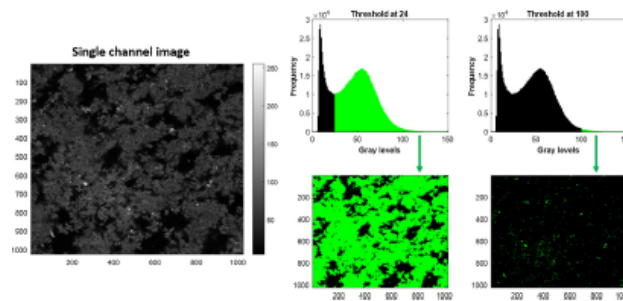
Existen en la actualidad diferentes formas que nos permiten clasificar las imágenes según seas sus características. Una de las formas más adecuadas es considerar la cantidad de información que un píxel puede contener. La información puede provenir de una simple diferencia en las intensidades con respecto a una referencia, generando un solo canal de información por píxel, por ejemplo, las imágenes de microscopía electrónica de barrido (SEM) o podemos encontrar imágenes que contienen más canales de información para cada píxel. Estas son las imágenes de espacio de color, las imágenes multi-espectrales y las imágenes hiper-espectrales [18].



**Figura 3.2.1 :** Imagen representativa de un píxel en una imagen.

### 3.2.1. Imagen de un Solo Canal

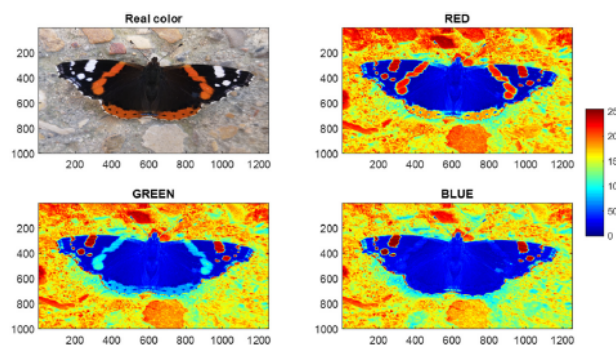
Las imágenes para las cuales hay un solo valor de intensidad para cada píxel se llaman imágenes mono-canal o imágenes de un solo canal y los valores que cada píxel contiene, puede varían según el tipo de dispositivo que toma la imagen, y normalmente se encuentran dentro de ciertos límites [18].



**Figura 3.2.1.1 :** Imagen de una muestra de yogurt obtenida mediante Microscopio Confocal Láser de Barrido (CLSM) [19].

### 3.2.2. Imágenes en Color

Las imágenes en espacio de color son las que intentan imitar la visión humana. En están compuestas normalmente por tres canales de información espectral, el rojo (alrededor de 700nm), el verde (alrededor de 550nm) y el azul (alrededor de 450nm), que combinados son capaces de recrear colores reales en lo que se conoce como el espacio de color RGB [18].

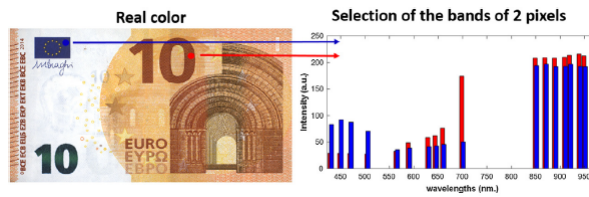


**Figura 3.2.2.1 :** Imagen en color real (RGB) de una mariposa y su representación en mono-canales rojo verde y azul [20].



### 3.2.3. Imágenes Multi-espectrales

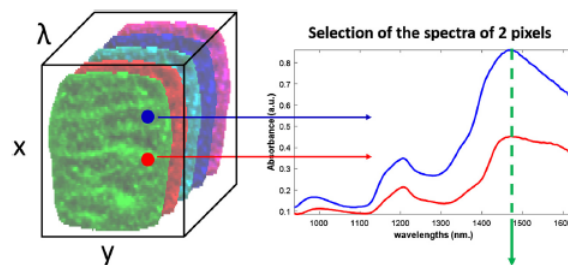
Las imágenes multi-banda o multi-espectrales son aquella que capturan imágenes individuales en longitudes de onda o longitudes específicas, frecuentemente obtenida mediante filtros o LED's específicos, a lo largo del espectro electromagnético (normalmente en las regiones visibles (VIS) e infrarrojo cercano (NIR)) [21]. en la parte izquierda de la figura 3.2.3.1 podemos observar una imagen RGB de un billete de 10 euros y en la parte derecha se pueden apreciar las intensidades de 18 diferentes longitudes de onda para dos de los píxeles.



**Figura 3.2.3.1** : Imagen multi-espectral tomada de un billete de 10 euros [18].

### 3.2.4. Imágenes Hiper-espectrales

Las imágenes hiper espectrales serán el tipo de imagen más relevante para el desarrollo de nuestro objetivo, puesto que la data contiene la respuesta espectral de cada muestra. Las imágenes hiper-espectrales son imágenes en las que se mide un espectro continuo para cada píxel [4] y se pueden obtener a partir de diferentes mediciones, entre las más populares son VIS, NIR, MIR por mencionar algunos. En la Figura 3.2.4.1 podemos apreciar como se ve una imagen hiper-espectral en el rango 940-1600nm y a la derecha se muestra la respuesta espectral de dos píxeles.



**Figura 3.2.4.1** : Representación de la imagen de una galleta medida con una cámara hiper-espectral.[18].

### 3.3. Tipos de Escaneo de Imágenes Hiper-espectrales

Existen tres enfoques básicos para la obtención de datos para imágenes hiper-espectrales. Dos configuraciones implican el desplazamiento de la muestra: configuración de escaneo puntual y configuración de escaneo en línea, (también conocidas como configuraciones de "whisker-broom" y "push-broom"), basadas en un modo de medición por paso y adquisición, donde se obtiene un solo espectro (escaneo puntual) o una línea de espectros (escaneo en línea) de la superficie en cada movimiento. El tercer enfoque es el sistema de escaneo plano, que no involucra ninguna parte móvil de la muestra, en esta configuración, la luz reflejada desde toda la superficie de la muestra pasa a través de filtros específicos [22]. La Figura 3.3.1 no muestra una representación más clara de lo señalado en este apartado.

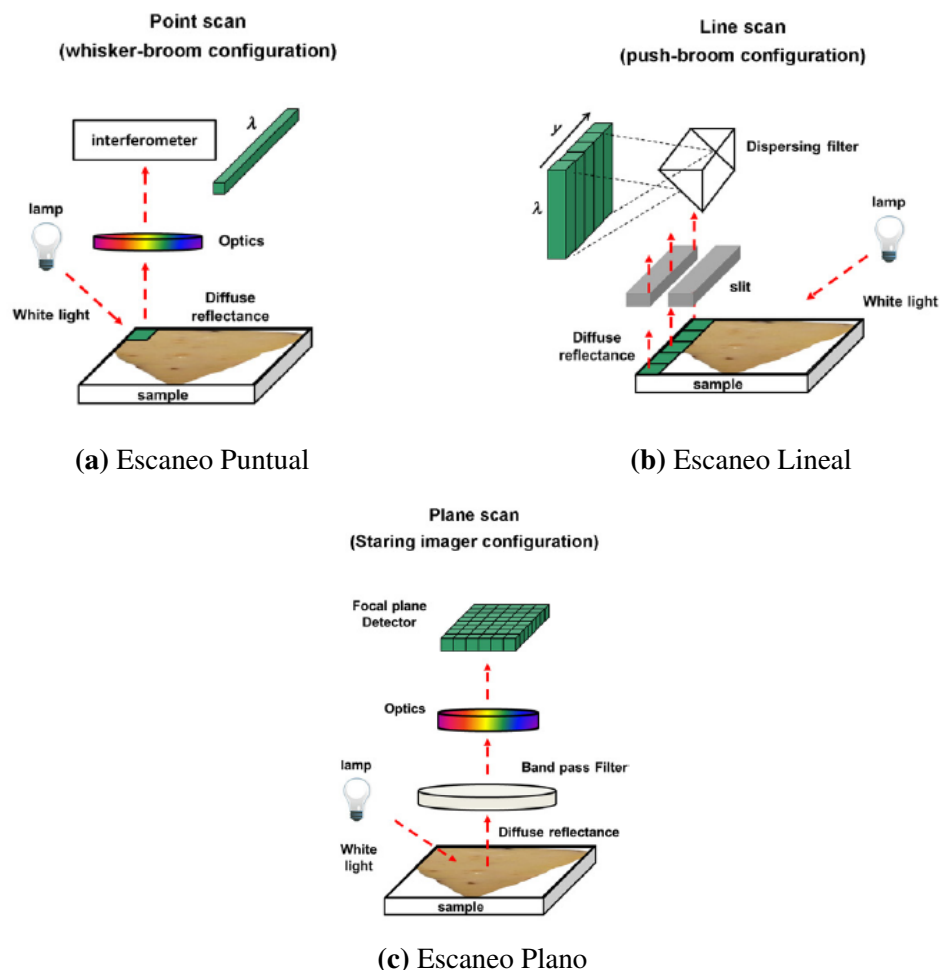
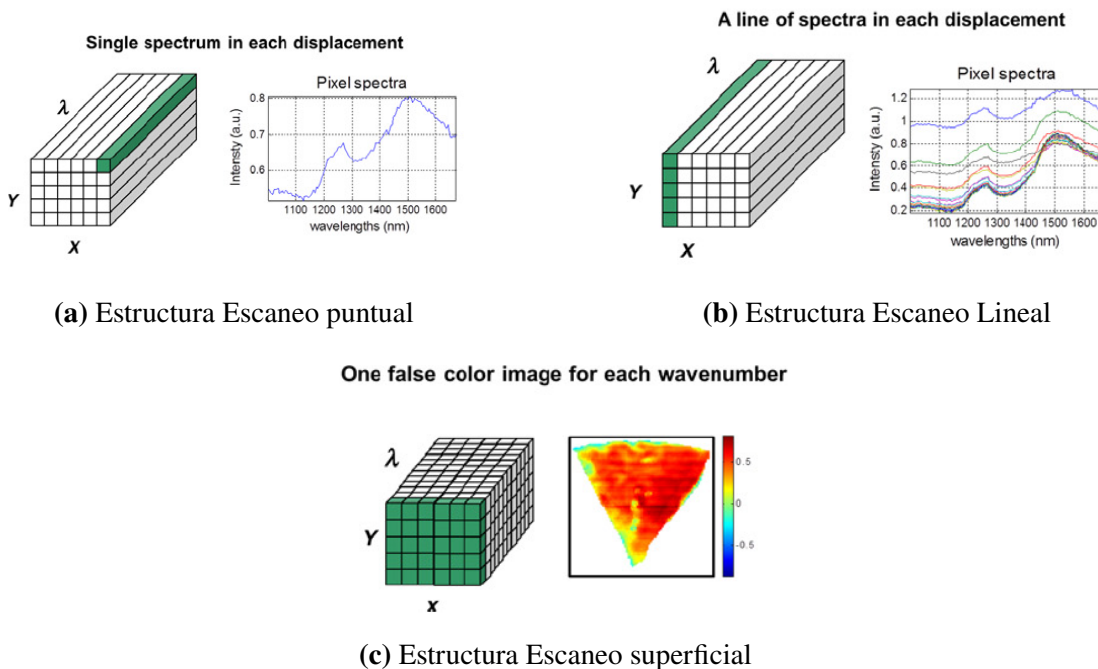


Figura 3.3.1 : Representación esquemática de las tres configuraciones comunes

### 3.4. Estructura de una Imagen Hiper-espectral

Como se mencionó existen diferentes formas y técnicas de adquirir los datos de una imagen hiper-espectral, pero el resultado final de la medición con cualquiera de los tres dispositivos es una estructura de datos tridimensional o un cubo de datos hiper-espectrales (hipercubo). Esta estructura también puede visualizarse como una matriz tridimensional  $\mathbf{D}$  de dimensiones  $(X \times Y \times \lambda)$ , en la que cada píxel (coordenada X-Y) es un espectro de  $\lambda$  longitudes de onda. [22].

Existen diferencias en la calidad de los datos recopilados y la utilidad de cada una de las configuraciones debe considerarse de antemano [23], [24]. Si se requiere una estructura más detallada es preferente ocupar una configuración de escaneo puntual, pero esta conllevará a un proceso considerablemente más lento para la adquisición de una imagen completa, lo que no permite mediciones en tiempo real requeridos en ciertos procesos. Para este tipo de proceso en los que se necesita un muestreo constante, es recomendable un escaneo plano, pero, para la obtención de imágenes de buena calidad es necesario optimizar el sistema. Estas consideraciones están fuera del alcance de este proyecto pero una discusión más detallada se pueden encontrar en las fuentes [25], [26].



**Figura 3.4.1 :** Representación esquemática de la estructura del hiper-cubo en los tres métodos

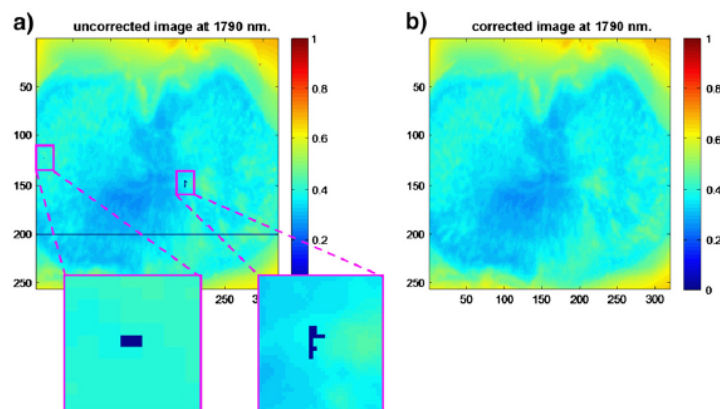
### 3.5. Pre-procesamiento de Imágenes Hiper-espectrales

Una de las principales características que posee una imagen hiper-espectral, es la gran cantidad de datos que se pueden almacenar en un píxel, esto sumado a la dimensión que pueda tener la imagen de interés nos puede llevar a una data compuesta de miles o millones de datos. Para manejar esta información y extraer la información relevante existen técnicas como análisis de componente principal [24] o la resolución de curvas multivariantes [27] que tienen gran éxito en la extracción de información deseada.

Sin embargo, existe un problema que no muchas veces es considerado pero merece ser tratado al trabajar con este tipo de imágenes y este corresponde a la presencia de valores erróneos en la data provocados por píxeles muertos, una distorsión provocada por el fondo de la imagen u observaciones atípicas que provocan inconsistencias en el modelo obtenido como resultado. Es por esto que realizar un buen pre-procesamiento de la imagen es de suma importancia si se requiere un modelo estimativo de alta exactitud. Estas observaciones anómalas pueden ser generadas por diversos factores [21].

#### 3.5.1. Eliminación de Píxeles Muertos

En general, aproximadamente el 1 % de los píxeles en los detectores NIR están “muertos” es decir, no contienen información útil [28]. Una de los métodos para reemplazar estos píxeles “muerto” es la técnica de interpolación con píxeles vecinos, ya que están fuertemente correlacionados, se realiza calculando la mediana o la media de los valores vecinos. Siendo la mediana la más favorable puesto que los valores extremos no le impactan de manera significativa [25].



**Figura 3.5.1.1 :** Extracción de píxeles muertos de una muestra [25]

### 3.5.2. Eliminación de Spikes

Los spikes son aumentos o disminuciones repentinas y agudas en el espectro [29], por lo que son fáciles de distinguir. Son generalmente causados por comportamientos anormales del dispositivo detector o debido a la saturación de luz en ciertas regiones espectrales. Aunque existen muchas alternativas en la literatura, la forma más directa de reemplazar los spikes es sustituirlos por la media o mediana del espectro de los píxeles vecinos [30]. En la Figura 3.5.2.1 se puede apreciar los resultados de realizar este cambio.

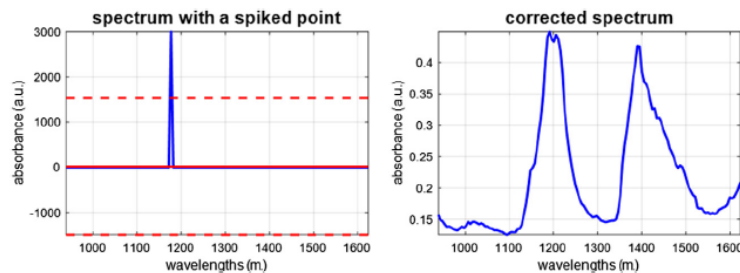


Figura 3.5.2.1 : Ejemplo de eliminación de spikes [30]

### 3.5.3. Eliminación de Fondo

La selección de la región de interés es un juego importante cuando la geometría de la muestra no cubre toda el área escaneada, el área que queda fuera de la muestra y los bordes de esta, generalmente está compuesta por espectros altamente ruidosos, dificultando el buen rendimiento de los modelos posteriores [30].

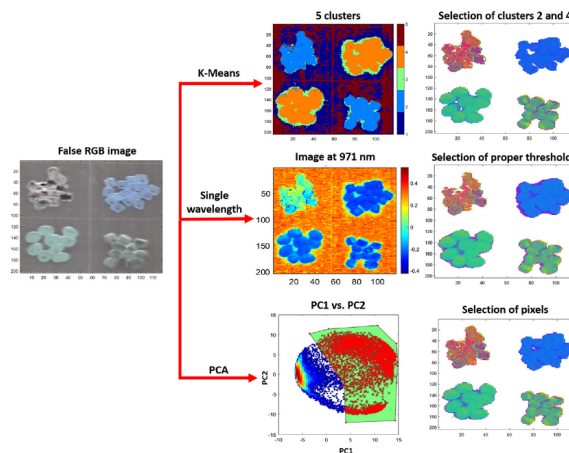


Figura 3.5.3.1 : 3 métodos de eliminación de fondo [30]

### 3.5.4. Pre-procesamiento Espectral

El objetivo del pre-procesamiento espectral es evitar las influencias de fenómenos indeseables que afecta la medición espectral, como la dispersión de la luz, efecto del tamaño de la partícula o diferencias morfológicas [24]. Algunas técnicas utilizadas para pre-procesamiento espectral son [25]: Eliminación de ruido, corrección de dispersión y derivada espectral. En la Figura 3.5.4.1 se puede observar los resultados obtenidos al aplicar los métodos mencionados.

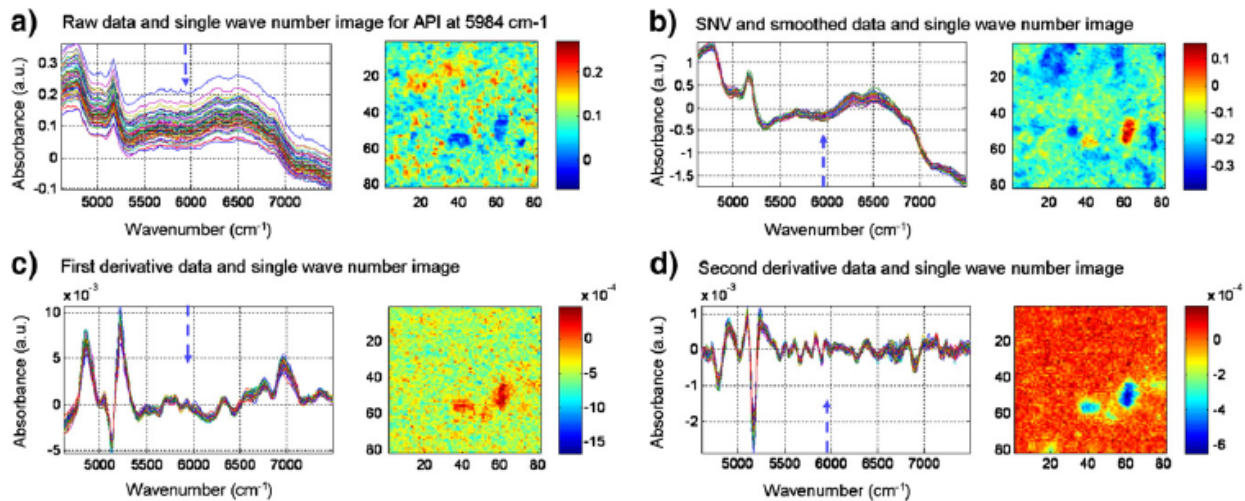
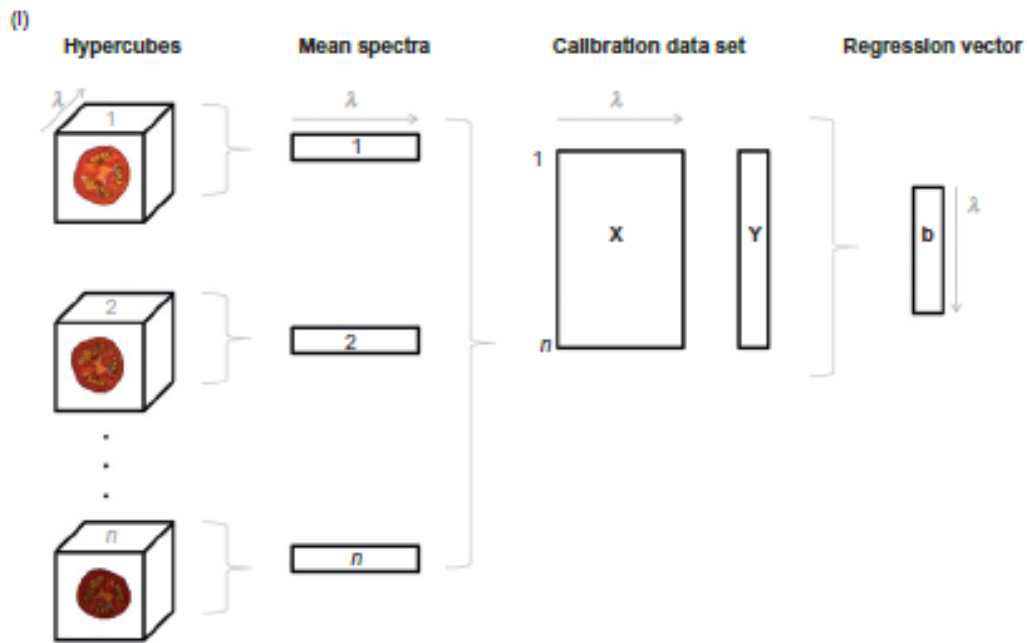


Figura 3.5.4.1 : Métodos pre-procesamiento espectral [25]

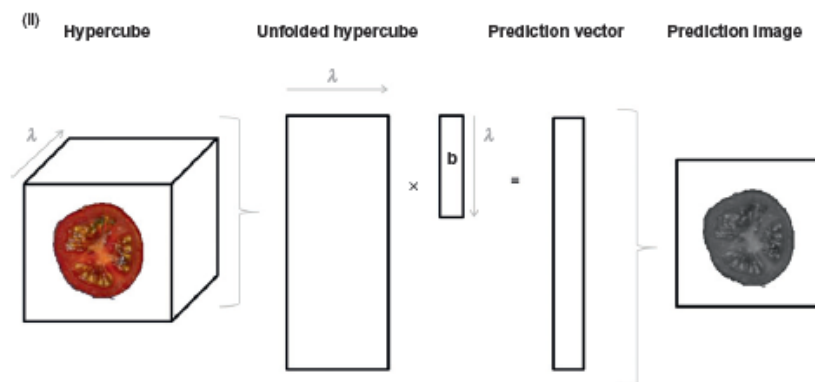
### 3.6. Obtención de Información Cuantitativa de Imágenes Hiper-espectrales Mediante Modelos de Regresión

Las imágenes hiper-espectrales al contener miles de datos por muestra podrían generar una problemática a la hora de procesar los datos. El enfoque típico para lidiar con esta problemática comúnmente consta de calcular un espectro promedio de la muestra que ha sido capturada, luego los espectros promedios de cada muestra se combinan para formar una  $\mathbf{X}$  a la que se le aplica un método de regresión (por ejemplo, PLS) a esta matriz para predecir los valores de referencia almacenados en la matriz  $\mathbf{Y}$  [22]. El proceso se resumen en la Figura 3.6.1 a continuación:



**Figura 3.6.1 :** Aplicación del modelado de regresión a imágenes hiper espectrales [22]

El vector  $b$  de la Figura 3.6.1 de regresión puede aplicarse a la imagen hiper-espectral original desplegada, dando como resultado los valores predicativos para píxel de la imagen. Al volver a plegar el vector “Prediction vector” de la figura 3.6.2 se obtiene una imagen de predicción [22], como se puede observar en la sección siguiente:



**Figura 3.6.2 :** Aplicación del modelo de regresión al hipercubo, resultando en una imagen de predicción [22].

### 3.7. Regresión por Mínimos Cuadrados Parciales (PLS)

Existen en la literatura una gran variedad de métodos de regresión. En específico para este estudio utilizaremos como base los métodos de regresión multivariada, que son aquellos donde se calcula la correlación entre un conjunto  $\mathbf{X}$  de  $(M \times \lambda)$  con  $M$  muestras independientes medidas en  $\lambda$  variables y una Variable predictora  $\mathbf{y}$  de dimensiones  $(M \times 1)$  [31], como sigue:

$$y = Xb + b_0 + e \quad (1)$$

Donde  $\mathbf{b}$  corresponde al vector de regresión que se quiere calcular y  $\mathbf{e}$  es el error asociado. Existen múltiples maneras de abordar este problema como por ejemplo método de mínimos cuadrados, regresión multilínea, regresión de componente principal por mencionar algunos pero en esta investigación nos centraremos en el método denominado mínimos cuadrados parciales (PLS) [32].

PLS es un método que considera tanto la varianza de  $\mathbf{X}$  como la covarianza de  $\mathbf{X}$  e  $\mathbf{Y}$ , así PLS proyecta  $\mathbf{X}$  e  $\mathbf{Y}$  en un nuevo espacio:

$$X = TP^T + E_X \quad (2)$$

$$Y = UQ^T + E_Y \quad (3)$$

Donde  $\mathbf{T}$  y  $\mathbf{U}$  son las matrices de scores de  $\mathbf{X}$  e  $\mathbf{Y}$  respectivamente y  $\mathbf{P}$  y  $\mathbf{Q}$  son sus matrices loadings asociadas. Las proyecciones mostradas en la ecuación anterior se realizan simultáneamente al maximizar la correlación entre  $\mathbf{T}$  y  $\mathbf{Q}$ :

$$U = RT \quad (4)$$

Así el vector de regresión puede ser calculado como se muestra en la ecuación (1). Según la cantidad de columnas que tenga el vector  $\mathbf{Y}$  es común referirse al método como PLS1 cuando solo existe una columna o PLS2 cuando  $\mathbf{Y}$  contiene dos o más columnas. PLS es un método muy utilizado y en la literatura podemos encontrar variadas aplicaciones.



## 4. Desarrollo

---

### 4.1. Introducción

En esta sección se presenta una exposición detallada de lo que se llevó a cabo durante el transcurso de la investigación dando cuenta del proceso de exploración, análisis y construcción del modelo de regresión por mínimos cuadrados parciales conocido como PLS.

### 4.2. Datos

Los datos utilizados en esta investigación corresponden a imágenes Hiper-espectrales tomadas a muestras de concentrados de cobre. En específico el conjunto corresponde a 32 archivos que contiene la respuesta espectral de cada una de las muestras y se proporciona además el porcentaje de la concentración mineralógica contenida en cada una de ellas.

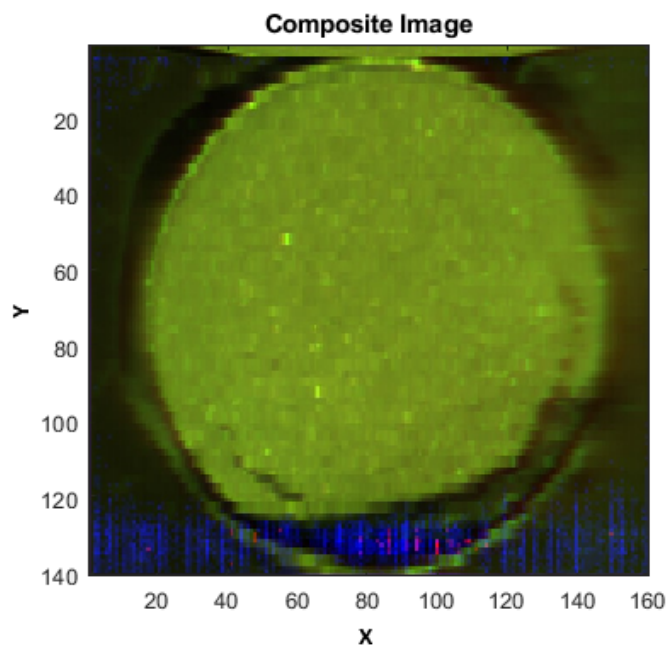
Como se ha mencionado las imágenes Hiper-espectrales corresponde a un tipo de imagen que captura información a lo largo de una banda espectral y puede ser representada como una matriz de 3 dimensiones  $(X, Y, \lambda)$ . En nuestro caso  $\lambda$  contiene 255 bandas en longitudes de onda que van desde los  $900nm$  a los  $1700nm$ ,  $X$  tendrá un largo de 160 y finalmente  $Y$  tiene un largo de 140. Así cada píxel puede ser modelado de la forma:

$$P(X, Y) = (p_1, \dots, p_\lambda) \quad (5)$$

Como se expuso en la sección 3.6, el método utilizado para estudiar los datos hiper-espectrales no aprovecha la información espacial de los píxeles contenidos en las imágenes, sino que, aprovecha solo el perfil espectral mediante la obtención de los promedio de respuesta espectral de cada banda. Es por esto que, aunque inicialmente debe reconstruirse el hipercubo, finalmente los datos serán tratados como un vector unidimensional de dimensiones  $(1, \lambda)$ , donde la medición  $\lambda_i$  está asociado al promedio de la  $i$ -ésima banda de longitud de onda.

### 4.3. Lectura de Datos

Lo primero que se debe realizar es cargar la base de datos con la que trabajaremos en Matlab y Hypertools. Los datos proporcionados como se mencionó corresponden imágenes hiper-espectrales tomadas de muestras de concentrados de cobre, estos datos están contenidos en archivos de extensión “.bin”, los que corresponde a archivos binarios, este tipo de archivos no pueden ser leídos de manera directa por la herramienta Hypertools, por lo que deben ser interpretados y transformados antes de poder utilizarlos. Para lograr esto nos ayudamos de la función “*fread*” para leer los datos y la función “*reshape*” para reconstruir y dar la forma adecuada a los datos y convertirlos en primera instancia, en una matriz de 3 dimensión (X , Y ,  $\lambda$ ), en donde X e Y corresponde al número de píxeles que contiene la imagen, y  $\lambda$  contiene la respuesta espectral de cada muestra.

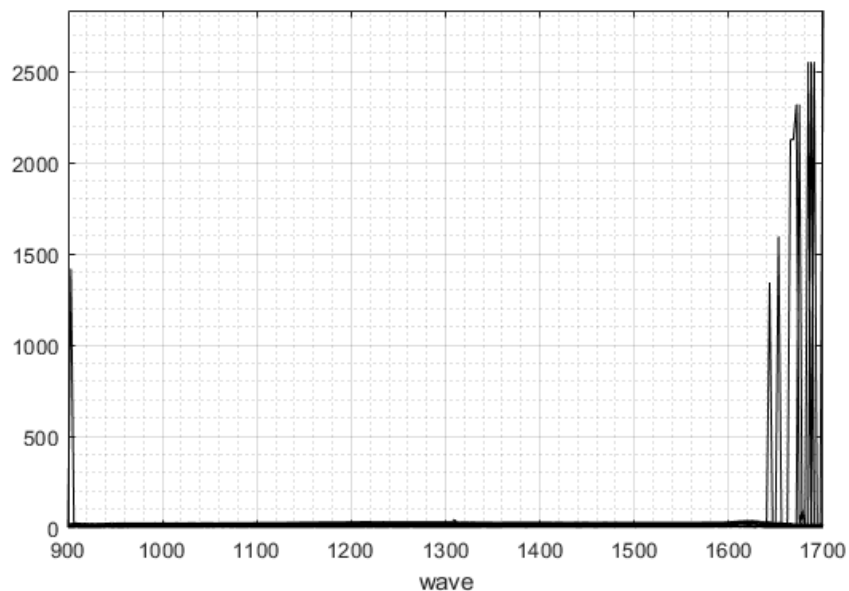


**Figura 4.3.1 :** Ejemplo de visualización de imagen (Muestra B002).

Una vez cargada y reconstruida la imagen en MATLAB, esta si puede ser visualizada directamente en la herramienta Hypertools como se ilustra en la Figura 4.3.1 . La imagen de la muestra que se aprecia en esta representación corresponde a una imagen compuesta en falso RGB y nos ayuda a analizar de manera visual el contenido intrínseco de cada una de las muestras capturadas ya que nos permite identificar características y patrones que no podrían ser evidentes al examinar los espectros de manera individual.

#### 4.4. Eliminación de Píxeles Muertos y Spikes

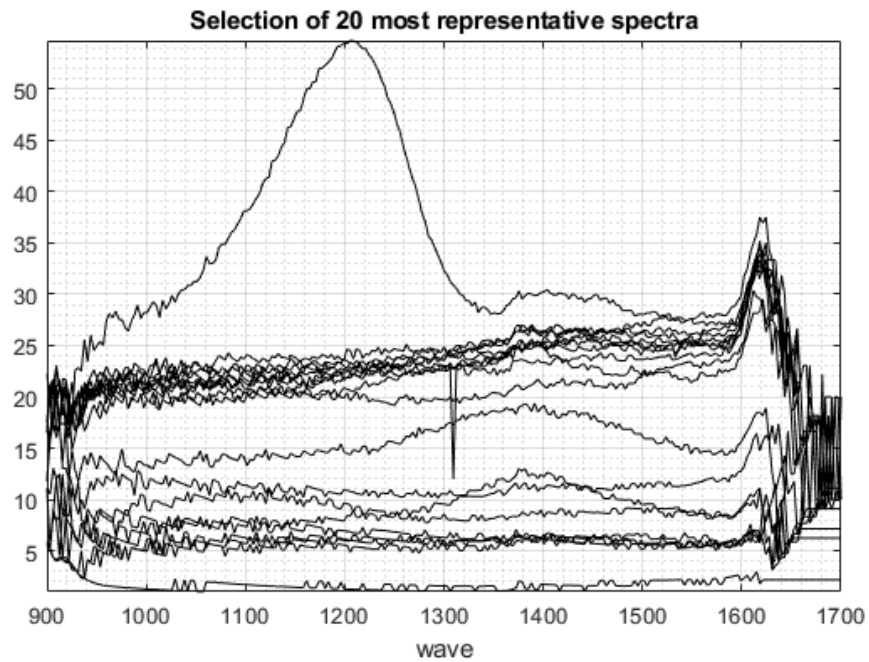
En la sección anterior se obtuvo un primer acercamiento sobre cómo se ve la data, de la Figura 4.3.1 es notorio observar en primer lugar una superficie irregular y destaca la presencia de píxeles con tonalidades azules y rojas que escapan de la tonalidad más regular de la imagen, esto nos evidencia que existe la presencia de píxeles “muertos” y/o spikes. En la figura 4.4.1 es posible apreciar que la respuesta de los 20 espectros más representativos de la muestra B002 en donde se observa que existen valores que alcanzan niveles excesivamente altos, lo que corrobora la presencia de estos.



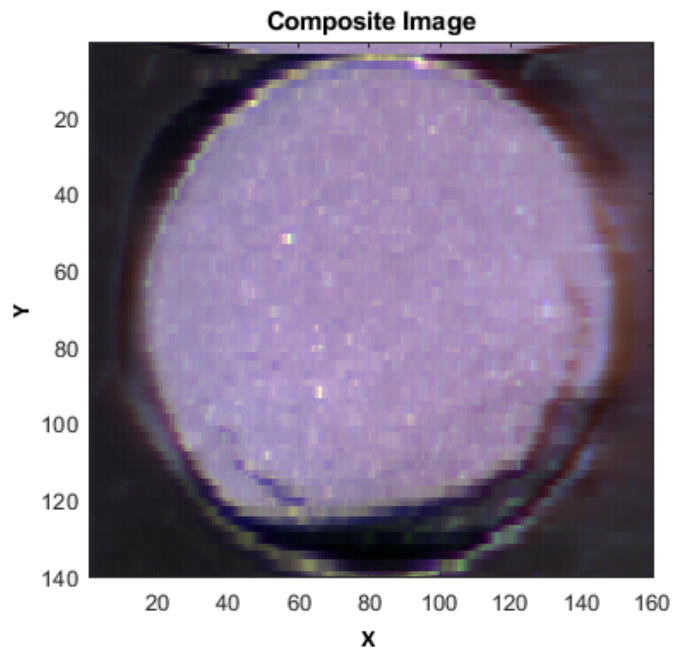
**Figura 4.4.1** : 20 espectros más representativos de muestra B002

La manera de eliminar estos datos que provocan distorsión es utilizando técnicas de interpolación, con la cual estos puntos son reemplazados por valores calculados basados en los puntos circundantes, lo que nos permitirá una visualización más precisa y continua de la información de cada píxel.

Los resultados obtenidos luego de aplicar esta técnica se muestran en la Figura 4.4.2 , en donde vemos que se eliminaron de manera efectiva los puntos donde de conflicto, tanto spikes como píxeles muertos y es posible apreciar de manera más clara y precisa el comportamiento de los 20 espectros más representativos de la muestra. En la Figura 4.4.3 podemos analizar estos cambios mencionados de manera visual, notando la ausencia de los píxeles con tonalidades rojas y azules y una tonalidad más homogénea de la imagen reconstituida en falso RGB.



**Figura 4.4.2 :** 20 Espectros mas representativos de Muestra B002 pre-procesada

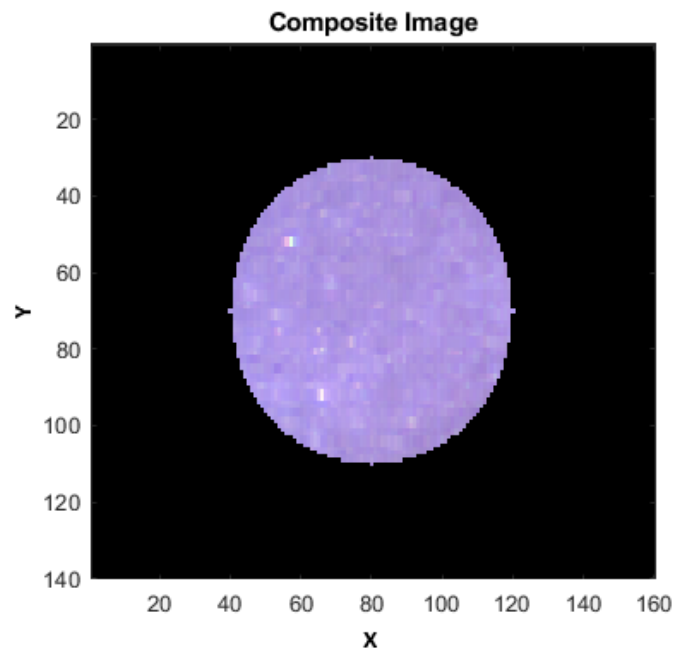


**Figura 4.4.3 :** Muestra B002 pre-procesada

Otro aspecto que es evidente notar a partir de la Figura 4.4.3 , es que las muestras poseen una morfología irregular y es claro que se debe seleccionar un área de trabajo adecuada que permita eliminar la información innecesaria para obtener resultados finales más precisos y confiables.

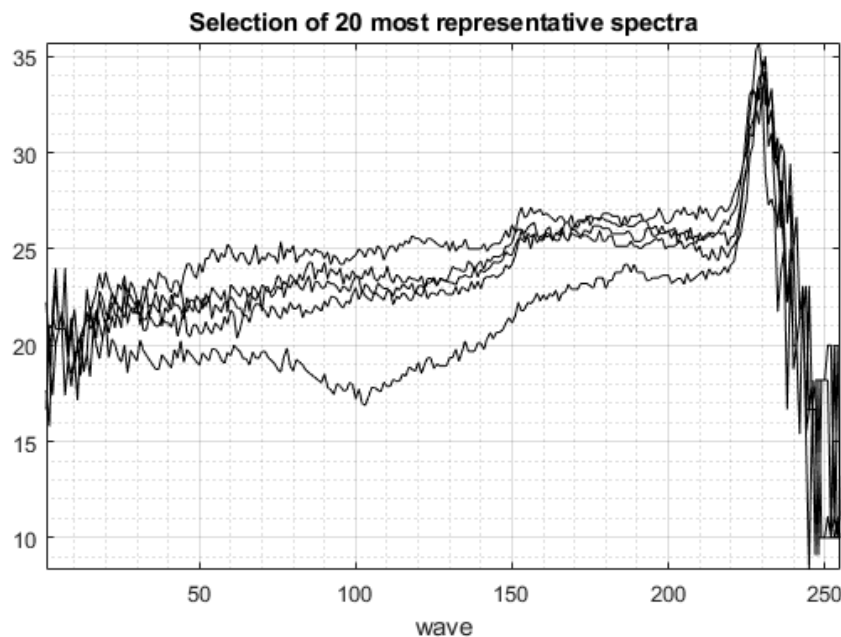
#### 4.5. Determinación del Área de Trabajo

Un aspecto importante a la hora de trabajar con imágenes hiper-espectrales tienes que ver con la zona de la imagen con la cual se quiere trabajar, esto debido a que la mayoría de las muestras al ser capturas incluyen la respuesta espectral del fondo lo cual se considera información innecesaria, en nuestro caso la información relevante de nuestras muestras se encuentra contenida en la circunferencia central, por lo que nos enfocaremos en esa zona. Para conseguir esto aplicaremos una máscara a cada una de las muestras y eliminaremos toda la zona externa a la circunferencia y también los bordes de esta, ya que estas zonas son las mayores causantes de distorsiones en la respuesta espectral por lo que nuestro dataset no sería afectado por estos factores.



**Figura 4.5.1 :** Determinación del área de trabajo (Muestra B002)

Como se observa en la Figura 4.5.1 se eliminaron la zonas mencionadas, quedando solo la zona central, la zona exterior no contiene información por lo que hypertools al crear la imagen compuesta en falso RGB, le asigna el color negro, pero no existe una respuesta espectral en esa zona, Además al realizar este cambio se logra observa de manera aún más clara el comportamiento de las respuestas espectrales de las muestras, ya que se eliminó la mayor parte del ruido existente en las muestras. Este cambio se puede apreciar en la Figura 4.5.2 , en donde se exhiben los espectros más representativos de la muestra B002.



**Figura 4.5.2 :** Espectros mas representativos (Muestra B002), luego se seleccionar área de trabajo

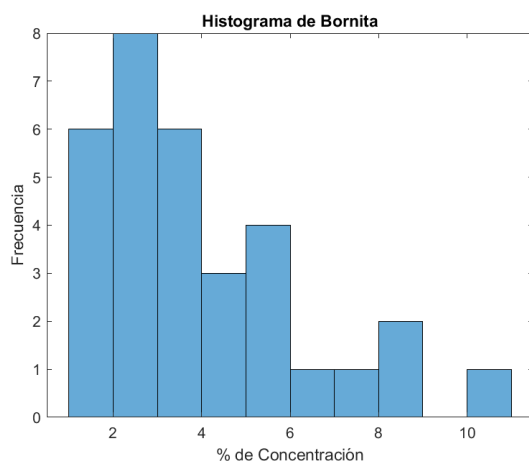
#### **4.6. Construcción matriz X e Y**

Para la creación de la matriz **X** se siguió el procedimiento mostrado en la Figura 3.6.1 , Como nuestra data esta conformada por por 32 muestras de concentrados de cobre la matriz tendrá dimensiones (32 x 255). En tanto, vector **Y** corresponderá a la concentración de las muestras, estas se encuentran contenidas dentro de un archivo Excel del cual fueron tomadas y se creo un vector **Y** para cada uno de los minerales contenidos en las muestras, en este caso el vector tendrá dimensiones (32 x 1). El código con el cual se construyeron estas matrices se puede analizar con mayor detalle en el Anexo A.

#### **4.7. Análisis de la Distribución Mineral de Concentrados de Cobre**

Uno de los factores de relevancia al momento de trabajar con concentrados de cobres es el análisis de la distribución mineral que estos poseen. Una de las técnicas de análisis que se utilizan, es mediante el uso de histogramas, debido a que estos nos entregan una representación visual que nos permite comprender rápidamente la distribución de los valores en un conjunto de datos.

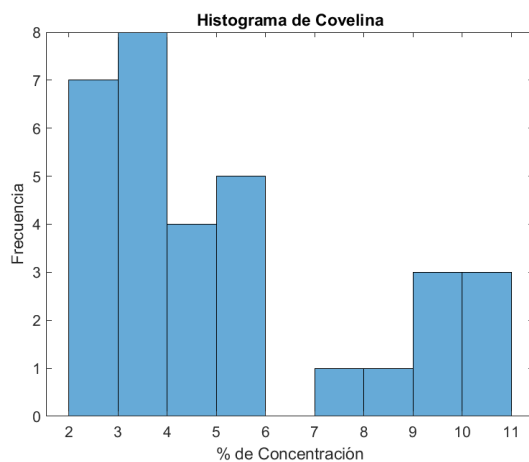
#### 4.7.1. Bornita



**Figura 4.7.1.1 : Histograma de Bornita**

Como se muestra en la Figura 4.7.1.1 , en el caso de la Bornita notamos que presenta un porcentaje de concentración que en su mayoría se mantienen bajo el 6 %, siendo el rango de 2 a 3 % donde se concentra el porcentaje de bornita más repetitivo de nuestro conjunto de datos.

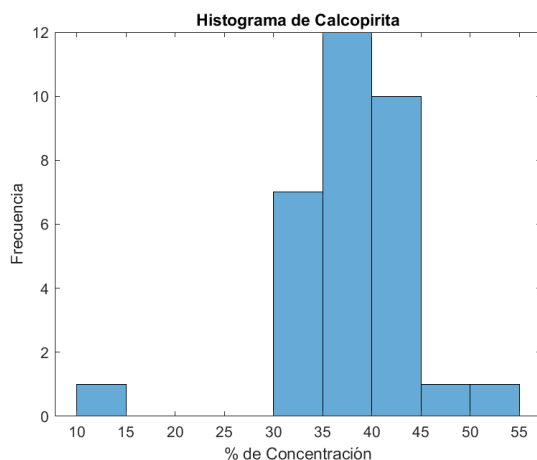
#### 4.7.2. Covelina



**Figura 4.7.2.1 : Histograma de Covelina**

Para la Covelina como se observa en la Figura 4.7.2.1 las cantidades de concentración de este mineral en las muestras tienden a posicionarse bajo el 6 %, pero además se logra observa una tendencia menor a niveles de concentración contenidos en el rango de 9-11 %.

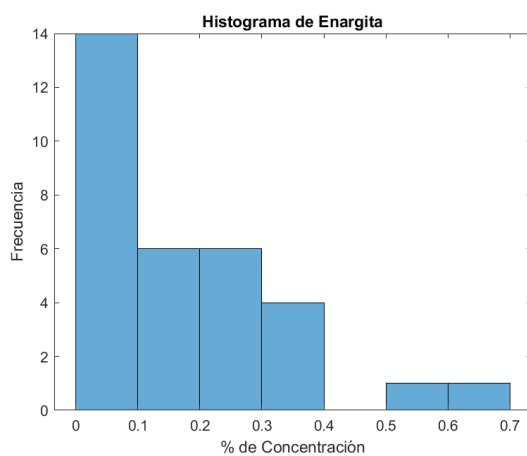
### 4.7.3. Calcopirita



**Figura 4.7.3.1 : Histograma de Calcopirita**

Para el caso la Calcopirita que se muestra en la Figura 4.7.3.1 se tiene que esta presenta niveles de concentración por sobre el 30 %, siendo el rango de más frecuencia de 35-40 % , Además se observa un caso aislado donde la concentración difiere de forma significativa del resto de muestras puesto que se encuentra bajo el 15 % lo que podría considerarse como un valor atípico.

### 4.7.4. Enargita

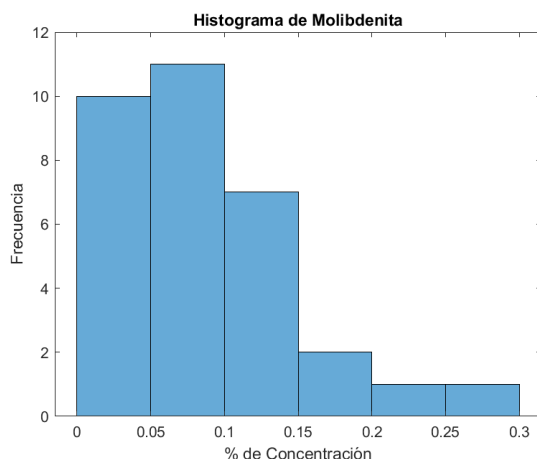


**Figura 4.7.4.1 : Histograma de Enargita**

En la Figura 4.7.4.1 , es claro notar que las concentraciones de enargita en las muestras son bajo el 1 %, la distribución nos muestra que existe cierta uniformidad de este mineral en las muestras y en su mayoría el porcentaje de este se encuentra en el rango de 0-0.1 %.



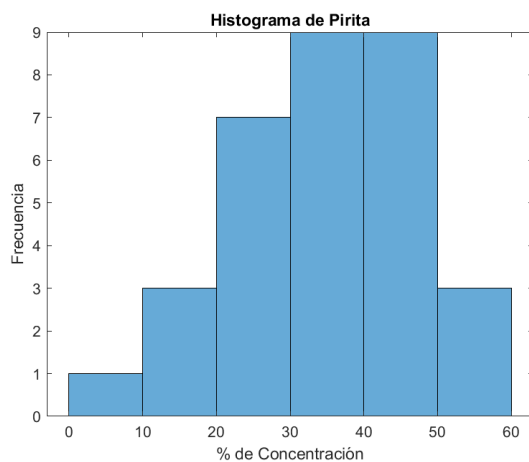
#### 4.7.5. Molibdenita



**Figura 4.7.5.1 : Histograma de Molibdenita**

La distribución concentración de Molibdenita (Figura 4.7.5.1 ) en el conjunto de datos, es principalmente baja manteniéndose en rangos que no superan el 0.3 %, en general la concentración de este mineral presenta mayor tendencia en el rango de 0.01 %.

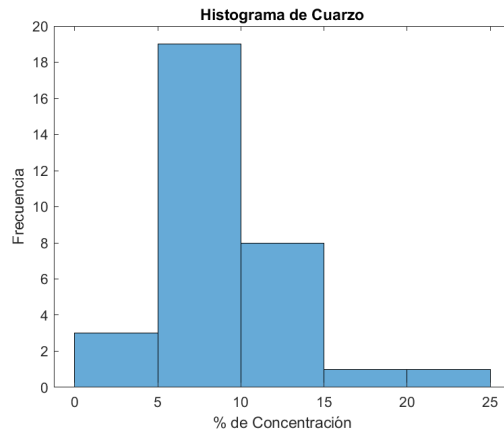
#### 4.7.6. Pirita



**Figura 4.7.6.1 : Histograma de Pirita**

En la Figura 4.7.6.1 donde se observa la distribución de concentración de la Pirita, se puede apreciar que este posee una alta variabilidad en las muestras que analizamos, además se observa del histograma que este sigue una distribución de tipo normal y su rango de mayor frecuencia se encuentra desde el 30-50 % de concentración.

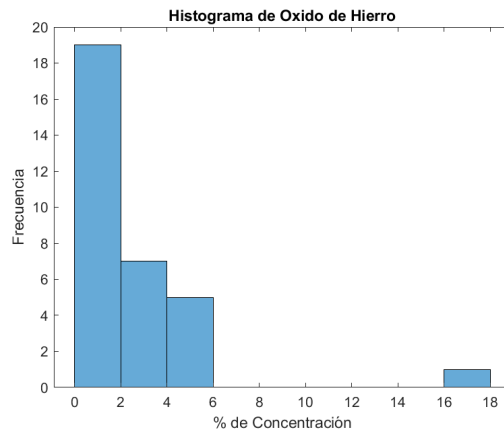
#### 4.7.7. Cuarzo



**Figura 4.7.7.1 : Histograma de Cuarzo**

Del mineral de cuarzo como se muestra en la Figura 4.7.7.1 , se puede observar que sus niveles de concentración se concentra en el rango de 5-15 %, siendo el rango de 5-10 % el de mayor frecuencia, además se aprecia que en casos de menor ocurrencia las muestras contienen niveles de porcentaje fuera de esta tendencia.

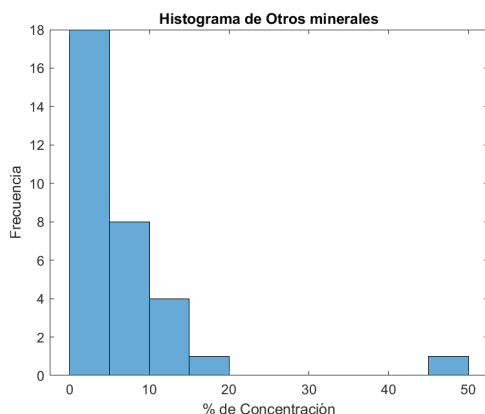
#### 4.7.8. Óxido de Hierro



**Figura 4.7.8.1 : Histograma de Óxido de Hierro**

Las concentraciones de Óxido de Hierro como se muestra en la Figura 4.7.8.1 se presenta en concentraciones menores al 6 % siendo el rango de tendencia entre 0-2 %, aunque en general este mineral representa una concentración baja se puede observar que en situaciones excepcionales también puede contener niveles más elevados como se aprecia en el histograma.

#### 4.7.9. Otros Metales



**Figura 4.7.9.1 :** Histograma de Otros Metales

Como se puede apreciar en la Figura 4.7.9.1 la concentración de otros minerales no identificados puede llegar a abarcar una gran proporción del contenido mineralógica de las muestras, aunque en general se encuentra por valores bajo el 20 % , marcando su tendencia en valores dentro del rango de 0-5 %.

#### 4.8. Selección de Parámetro para el Modelo PLS

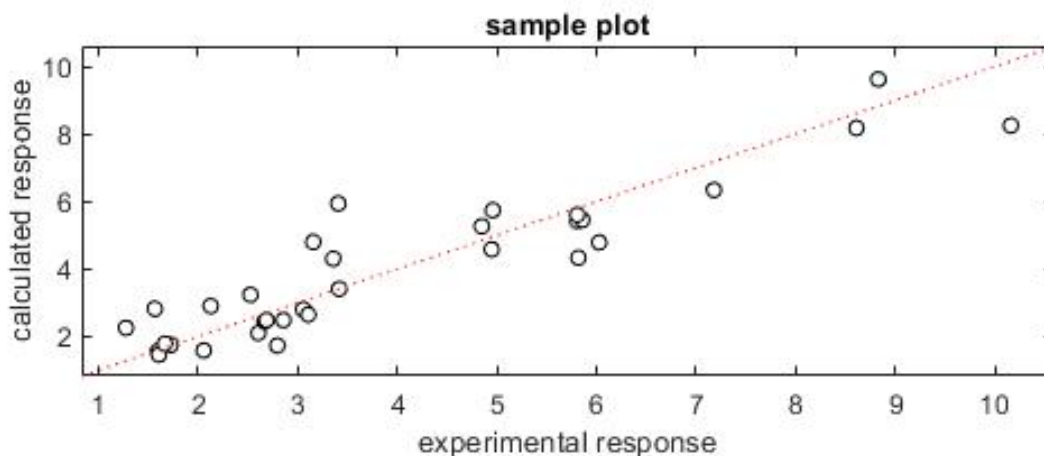
La aplicación del Modelo de PLS se realizará directamente dentro del herramienta Hypertools, para esto se necesitan las matrices **X** e **Y** ya obtenidas previamente, además de esto se debieron seleccionar opciones para aplicar el proceso de validación cruzada, el cual es necesario para entrenar y validar el modelo que se generado. las opciones escogidas son:

- **Mean centering:** Corresponde a una técnica de preprocesamiento que consiste en restar la media a cada variable a todos los valores de esa variable. Esto con el fin de evitar que las variables latentes sean dominadas por valores de variables en escalas diferentes.
- **Venetian Blinds:** Es una técnica de validación cruzada para evaluar el rendimiento de modelos predictivo. Su funcionamiento se basa en dividir los datos en k particiones. Para cada iteración se utiliza k-1 partición para entrenar el modelo y se usa la partición restante como conjunto de prueba.
- **Numero CV:** Corresponde al número de particiones que utilizara el método de validación cruzada.

## 5. Resultados

### 5.1. Modelo PLS para Bornita ( $\text{Cu}_5\text{FeS}_4$ )

Luego de aplicar el método de regresión PLS, se obtuvieron los resultados que se exponen a continuación:



**Figura 5.1.1 :** Modelo PLS Bornita

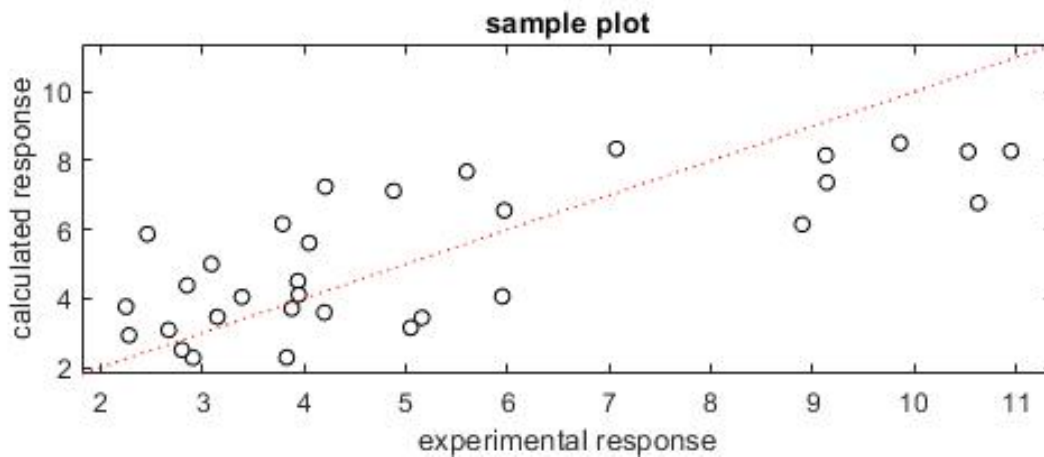
Métricas de rendimiento			
$R^2$	0.838	$R^2_{cv}$	0.465
RMSEC	0.924	$RMSEC_{cv}$	1.681

**Tabla 5.1.1 :** Rendimiento Modelo PLS Bornita

A partir de la Tabla 5.1.1 podemos apreciar la respuesta del modelo PLS en la que nos entrega un valor de  $R^2=0.838$ , que nos indica que el modelo posee un buen poder explicativo de la variabilidad de la concentración de Bornita. Además, señala que el modelo en promedio tiene un error de predicción de 0.924 unidades. Estos resultados se reflejan de manera gráfica en la Figura 5.1.1 en donde podemos notar que la línea roja tiende a seguir de manera cercano los valores de concentración experimentales. En el caso de las métricas de validación cruzada el valor de  $R^2_{cv}$  es de 0.465 lo que sugiere que posee una capacidad de generalización moderado a datos nuevos o no vistos y el promedio de error  $RMSEC_{cv}$  aumenta su valor a 1.681 unidades.

## 5.2. Modelo PLS para Covelina (CuS<sub>2</sub>)

Los resultados luego de la aplicación del método de regresión PLS son:



**Figura 5.2.1 :** Modelo PLS Covelina

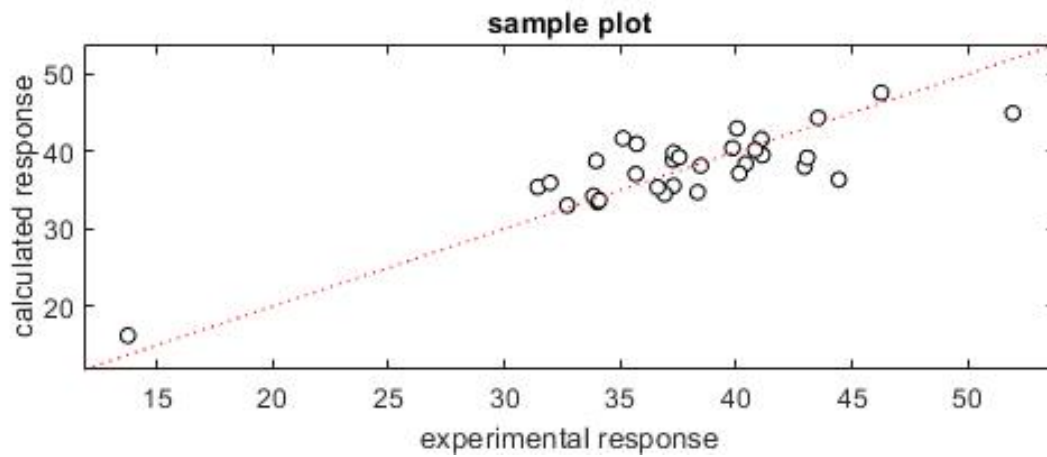
Métricas de rendimiento			
$R^2$	0.555	$R^2_{cv}$	0.017
RMSEC	1.803	$RMSEC_{cv}$	2.726

**Tabla 5.2.1 :** Rendimiento Modelo PLS Covelina

Para el caso de la Covelina en la Tabla 5.2.1 tenemos que el valor de  $R^2$  es de 0.555 lo que representa que el modelo explica el 55.5 % de la variabilidad en la respuesta. En cuanto al error promedio tenemos que RMSEC es de 1.803 unidades. Para el caso de las métricas en validación cruzada tenemos que el modelo nos entrega un valor de  $R^2_{cv}=0.017$ , lo que indica que este presenta dificultades para generalizar en nuevos datos y podría estar sobre ajustando los datos de entrenamiento. En la Figura 5.2.1 podemos apreciar que los datos se concentran mayoritariamente en una porción del gráfico y aunque el modelo intenta buscar una correlación que describa el comportamiento de la variable solo logra capturar una parte de ella.

### 5.3. Modelo PLS para Calcopirita (CuFeS<sub>2</sub>)

Para el modelo de regresión PLS para calcopirita se obtuvo lo siguiente:



**Figura 5.3.1 :** Modelo PLS Calcopirita

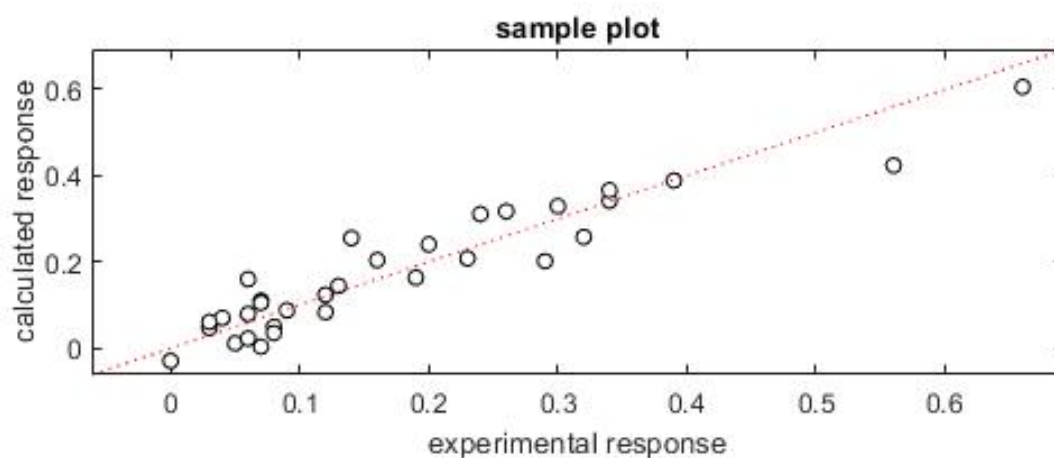
Métricas de rendimiento			
$R^2$	0.710	$R^2_{cv}$	0.216
RMSEC	3.312	$RMSEC_{cv}$	5.449

**Tabla 5.3.1 :** Rendimiento Modelo PLS Calcopirita

De la Figura 5.3.1 es posible apreciar que la línea de respuesta generada por el modelo logra representar de manera más eficiente la concentración mineralógica de las muestras. En la Tabla 5.3.1 podemos corroborar estos resultados, teniendo un  $R^2$  de 0.71, lo que se traduce que el modelo explica el 71 % de la variabilidad en la respuesta. El error promedio de cálculo es de  $RMSEC=3.312$  unidades. Para los datos de validación cruzada la variable  $R^2_{cv}$  es de 0.216 lo que señala que el modelo tiene dificultades para generalizar a datos nuevos, lo que se ve reflejado también en el promedio de error  $RMSEC_{cv}$  que corresponde a 5.449 unidades.

#### 5.4. Modelo PLS para Enargita ( $\text{Cu}_3\text{AsS}_4$ )

Los resultados del modelo de regresión para el mineral Enargita son:



**Figura 5.4.1 :** Modelo PLS Enargita

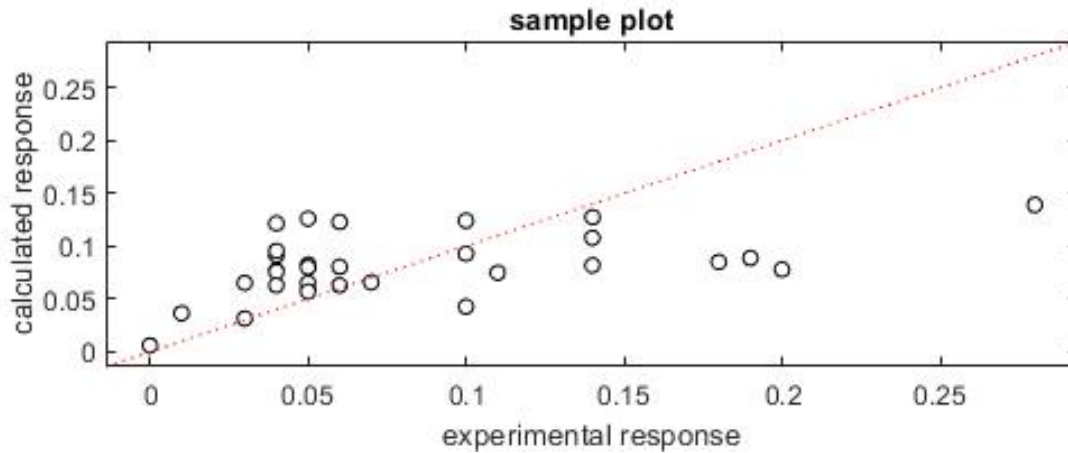
<b>Métricas de rendimiento</b>			
$R^2$	0.883	$R_{cv}^2$	0.482
RMSEC	0.053	$RMSEC_{cv}$	0.111

**Tabla 5.4.1 :** Rendimiento Modelo PLS Enargita

En la Figura 5.4.1 se revela que la recta obtenida del Modelo PLS para el mineral de enargita, presenta una respuesta con un alto nivel de concordancia con los puntos de dispersión del conjunto de datos analizado. Esto se ve reflejado en la Tabla 5.4.1 en donde la métrica de correlación  $R^2$  es de 0.883, por lo que el modelo puede explicar alrededor del 88.3 % de la variabilidad de las respuestas experimentales, con un error promedio RMSEC de 0.053 unidades. En el caso de la validación cruzada se tiene que  $R_{cv}^2$  es de 0.482, esto indica que el modelo tiene una capacidad de predicción aceptable en datos no vistos, con un promedio de error  $RMSEC_{cv}$  de 0.111 unidades.

## 5.5. Modelo PLS para Molibdenita (MoS<sub>2</sub>)

Para el mineral Molibdenita luego de aplicar el método de regresión PLS se obtuvo el siguiente comportamiento del modelo:



**Figura 5.5.1 :** Modelo PLS Molibdenita

Métricas de rendimiento			
$R^2$	0.228	$R^2_{cv}$	0.107
RMSEC	0.054	$RMSEC_{cv}$	0.059

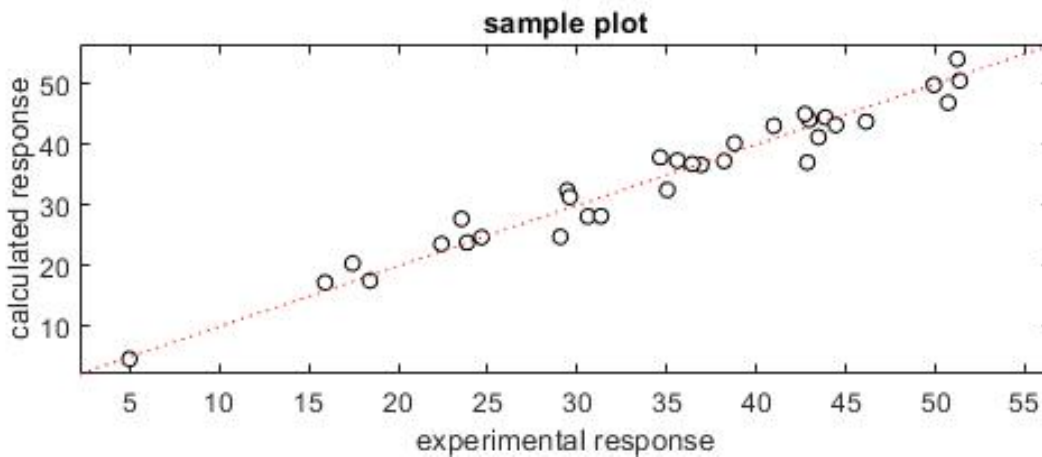
**Tabla 5.5.1 :** Rendimiento Modelo PLS Molibdenita

Es notorio apreciar de la Figura 5.5.1 que el modelo no logra adaptarse de manera adecuada a la variabilidad de los datos analizados. en la Tabla 5.5.1 queda en evidencia los bajos valores de  $R^2$  y  $R^2_{cv}$  que corresponde a 0.228 y 0.107 respectivamente, que ratifican el bajo porcentaje de eficiencia que el modelo presenta en este caso. Además, aunque los rango de error promedio de RMSEC y  $RMSEC_{cv}$ , corresponde a valores cercanos a cero (0.054 y 0.059 unidades), hay que tener en consideración que el porcentaje de concentración de este mineral se encuentra bajo el 0.3 % por lo que un error en del orden de 0.05 unidades podría inducir un error elevado al momento de realizar predicciones con el modelo.



## 5.6. Modelo PLS para Pirita (FeS<sub>2</sub>)

El comportamiento del modelo PLS para el mineral Pirita se muestra a continuación:



**Figura 5.6.1 :** Modelo PLS Pirita

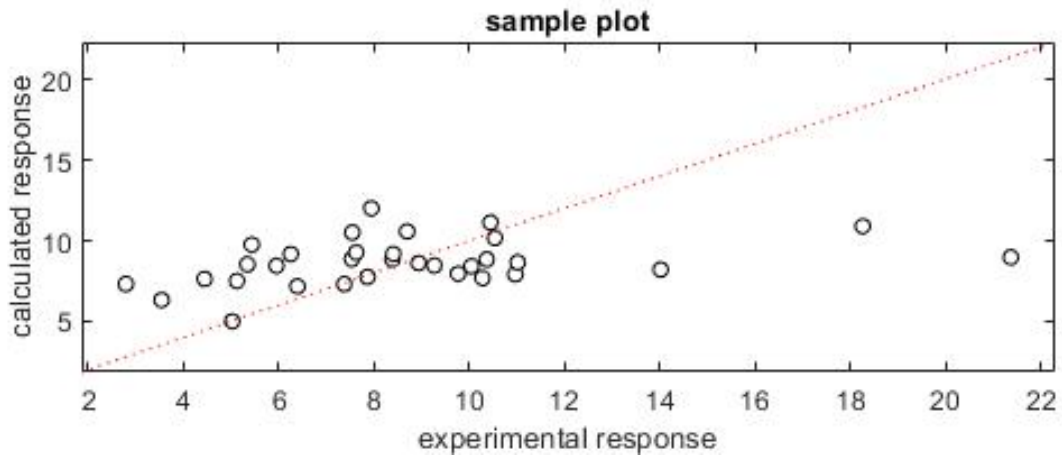
Métricas de rendimiento			
$R^2$	0.956	$R_{cv}^2$	0.641
RMSEC	2.374	$RMSEC_{cv}$	6.785

**Tabla 5.6.1 :** Rendimiento Modelo PLS Pirita

En la Figura 5.6.1 se observa que la relación lineal modelada exhibe un alto nivel de concordancia con los puntos de dispersión de concentración de Pirita de los datos analizados. La Tabla 5.6.1 expone la alta eficiencia que se logra en este caso, con valores de  $R^2$  de 0.956, el cual indica que posee una capacidad sólida para explicar la variabilidad de los datos con un error promedio de RMSEC de 2.374 unidades. Para el caso de las mediciones en validación cruzada tenemos que  $R_{cv}^2$  es de 0.641, lo que refleja una buena adaptación del modelo a datos no vistos con un error  $RMSEC_{cv}$  de 6.785 unidades.

### 5.7. Modelo PLS para Cuarzo (SiO<sub>2</sub>)

Los resultados obtenidos para el mineral de Cuarzo son lo siguientes:



**Figura 5.7.1 :** Modelo PLS Cuarzo

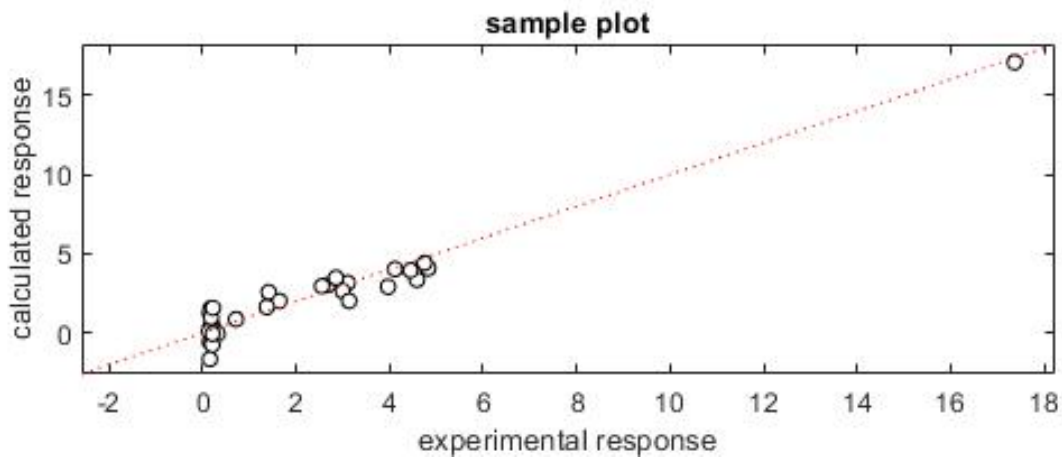
Métricas de rendimiento			
$R^2$	0.140	$R^2_{cv}$	0.100
RMSEC	3.513	$RMSEC_{cv}$	3.972

**Tabla 5.7.1 :** Rendimiento Modelo PLS Cuarzo

En el caso del mineral de cuarzo podemos analizar en la Figura 5.7.1 que el modelo aunque trata de encontrar una correlación entre las variables dependientes e independientes, este solo lo logra en una porción muy ínfima del comportamiento. En la Tabla 5.7.1 notamos que las métricas de  $R^2$  y  $R^2_{cv}$  tienen valores de 0.140 y 0.100 respectivamente evidenciando las dificultades del modelo para ajustarse. Bajo la misma línea los errores promedios presentan valores elevados, lo que confirma la ineficiencia que se obtuvo de este modelo en particular para el mineral de cuarzo.

## 5.8. Modelo PLS para Oxido de Hierro ( $Fe_xO_y$ )

Los resultados obtenidos al aplicar el modelo para el mineral de óxido de hierro son:



**Figura 5.8.1 :** Modelo PLS Oxido de Hierro

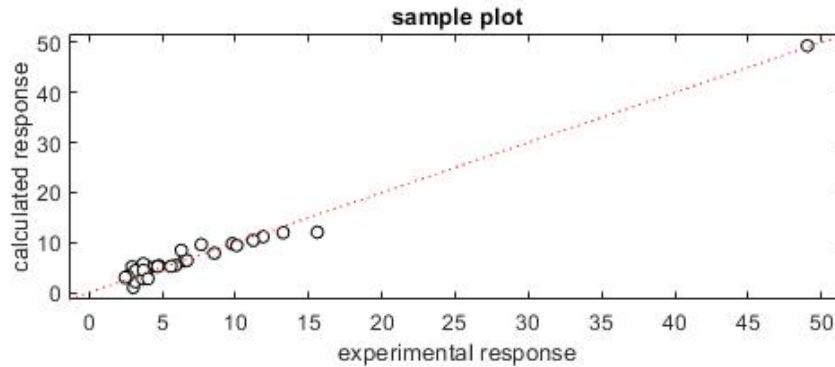
Métricas de rendimiento			
$R^2$	0.940	$R^2_{cv}$	0.629
$RMSEC$	0.780	$RMSEC_{cv}$	1.951

**Tabla 5.8.1 :** Rendimiento Modelo PLS Oxido de Hierro

En la figura 5.8.1 es posible apreciar que el modelo se comporta de manera altamente correlacionada con los valores de concentración de mineral de óxido de hierro, inclusive en puntos que escapan de la tendencia de concentración. De la Tabla 5.8.1 se tiene que  $R^2$  es igual a 0.940 lo cual indica que el modelo es capaz de representar la variabilidad de una manera óptima y con errores promedios de  $RMSEC$  de 0.780 unidades. Para las métricas de validación cruzada tenemos que  $R^2_{cv}$  es igual a 0.629, esto señala que el modelo es capaz de adaptarse de manera muy eficiente a datos no vistos con un error promedio  $RMSEC_{cv}$  de 1.951 unidades.

## 5.9. Modelo PLS para Otros Metales

El ultimo caso corresponde a la aplicación del método de regresión PLS a la concentración de muestras de otros metales presentes en los concentrados de cobre, su resultado es el siguiente:



**Figura 5.9.1 :** Modelo PLS Otros Metales

Métricas de rendimiento			
$R^2$	0.977	$R^2_{cv}$	0.417
RMSEC	1.234	RMSEC cv	6.27

**Tabla 5.9.1 :** Rendimiento Modelo PLS Otros Metales

De la Figura 5.9.1 podemos notar que la concentración de otros metales en los datos analizados tiende a concentrarse en la zona de menor porcentaje y la recta generada por el modelo tiene un gran capacidad de representar la tendencia de estos. En la Tabla5.9.1 se obtiene que el valor de  $R^2$  es de 0.977 lo que refleja una alta adaptación del modelo a la variabilidad de los datos de estudio con un error promedio RMSEC de 1.234 unidades. Para las métricas de validación cruzada tenemos que  $R^2_{cv}$  es igual a 0.417 lo que refleja una adecuada capacidad del modelo de generalizar para nuevos datos.

## 6. Conclusiones

---

### 6.1. Discusión

Es posible señalar que para el desarrollo de un modelo de regresión utilizando imágenes hiper-espectrales se debe comprender primeramente el tipo de archivo con el que se está trabajando, puesto que este tipo de imágenes son muy sensibles al ruido, por lo que analizar el comportamiento de la respuesta espectral y hacer un buen pre-procesamiento de la data es parte fundamental al trabajar con ellas en cualquier tipo de estudios.

Además se debe seleccionar de manera óptima los parámetros con los cuales se construirá el modelo de regresión, es importante manejar el concepto de validación cruzada, ya que la segmentación de nuestra data en zonas de entrenamiento y validación influyen en que nuestro modelo sea más o menos eficiente.

### 6.2. Conclusión

Como se revisó en la literatura existen diversidad de estudios en diferentes ramas de la ciencia, donde se utilizan imágenes hiper espectrales para la estimación de datos teniendo generalmente una alta eficiencia. En el estudio realizado en este documento los resultados fueron variados, pero general se obtuvo una buena representación por parte de los modelos lo que corrobora la eficiencia de la utilización del método de regresión PLS.

Dentro de los minerales que se analizaron en el apartado anterior, los que obtuvieron resultados que representan un mayor nivel de eficiencia corresponde a los minerales: Bornita, Calcopirita, Enargita, Pirita, Óxido de hierro y Otros metales. En ellos se observó que los modelos generaban una alta correlación y eran capaces de describir de manera eficiente la variabilidad de los porcentajes de concentración del respectivo mineral. Además de esto las métricas proporcionadas a partir de los modelos nos indicaban que estos poseen la capacidad de responder de una manera óptima al momento de predecir en datos que no se utilizaron en el entrenamiento. Estos resultados respaldan la eficacia del método de PLS y el porque es diversamente utilizado en estudios científicos es los que se necesitan realizar estimaciones.

Una respuesta más regular se obtuvo en caso de la Covelina, que logra una representatividad

de datos de 55.5 %, pero el modelo no presenta una buena respuesta a datos nuevos. Por otra parte, se tiene que, para los minerales de Molibdenita y Cuarzo, el modelo no logra encontrar una correlación que logre explicar de manera precisa y consistente la variabilidad de los datos teniendo una eficiencia bajo el 25 %. Los factores que podrían estar causando esto no siempre tiene que ver con la elección del modelo de regresión, factores externos como el bajo porcentaje de concentración del mineral sumado a la reducción del área de trabajo para eliminar ruido de las imágenes hiper-espectrales influir en que la respuesta espectral del mineral no será tan representativa a la hora de construir la matriz X de promedios, otro factor puede deberse a la distribución de las muestras en donde el modelo podría estar generando un sobre ajuste y por esto obtener resultados que lo hagan menos eficiente.

Finalmente se puede concluir que en términos generales se obtuvieron resultados satisfactorios en la creación de modelos utilizando la técnica de regresión PLS, evidenciando su alta capacidad de procesamiento y adaptación a conjuntos complejos de datos. Por otra parte, se debe destacar la versatilidad de la herramienta Hypertools que permite de una manera mucho más directa obtener este tipo de modelos matemáticos de estimación y presenta una alta variedad de otras técnicas y herramientas según las necesidades del estudio en cuestión.

### **6.3. Trabajo Futuro**

Para mejorar la exactitud con la cual predicen los modelos de regresión planteados se podrían evaluar las alternativas como:

- Trabajar con una data de mayor tamaño, al aumentar la cantidad de datos, es posible que se encuentre una mayor tasa de variabilidad en las concentraciones mineralogías de las muestras, por lo eventualmente el modelo encontraría una correlación mas adecuada, mejorando su precisión.
- Evaluar otras bandas de frecuencia, en el desarrollo de este trabajo se trabajo con HSI en el espectro NIR, por lo que estudiar el comportamiento espectral en otras bandas podría proporcionar mas información y eventualmente mejorar el modelo.
- Evaluar resultados en otro tipo de modelos, en esta investigación se centro en la utilización del modelo PLS que busca una relación lineal entre las variables dependientes e independientes.

Por lo que aplicar modelos mas complejos o no lineales para estudiar como estos capturan la relación entre variables podría ser otra alternativa.

## Referencias

---

- [1] G. ElMasry y D.-W. Sun, "CHAPTER 1 - Principles of Hyperspectral Imaging Technology", en *Hyperspectral Imaging for Food Quality Analysis and Control*, D.-W. Sun, ed., San Diego: Academic Press, 2010, págs. 3-43, ISBN: 978-0-12-374753-2. DOI: <https://doi.org/10.1016/B978-0-12-374753-2.10001-2>. dirección: <https://www.sciencedirect.com/science/article/pii/B9780123747532100012>.
- [2] J. Qin, M. Kim, K. Chao, D. Chan, S. Delwiche y B.-K. Cho, "Line-Scan Hyperspectral Imaging Techniques for Food Safety and Quality Applications", *Applied Sciences*, vol. 7, pág. 125, ene. de 2017. DOI: [10.3390/app7020125](https://doi.org/10.3390/app7020125).
- [3] A. Roman-Gonzalez y N. I. Vargas-Cuentas, "Análisis de imágenes hiperespectrales", *Revista Ingeniería & Desarrollo*, vol. 9, n.º 35, págs. 14-17, 2013.
- [4] J. Amigo, H. Babamoradi y S. Elcoroaristizabal, "Hyperspectral image analysis. A tutorial", *Analytica Chimica Acta*, vol. 896, págs. 34-51, 2015. DOI: [10.1016/j.aca.2015.09.030](https://doi.org/10.1016/j.aca.2015.09.030). dirección: <https://doi.org/10.1016/j.aca.2015.09.030>.
- [5] Å. Rinnan, F. van den Berg y S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra", *TrAC Trends in Analytical Chemistry*, vol. 28, n.º 10, págs. 1201-1222, 2009, ISSN: 0165-9936. DOI: <https://doi.org/10.1016/j.trac.2009.07.007>. dirección: <https://www.sciencedirect.com/science/article/pii/S0165993609001629>.
- [6] J.-C. Boulet y J.-M. Roger, "Pretreatments by means of orthogonal projections", *Chemometrics and Intelligent Laboratory Systems*, vol. 117, págs. 61-69, 2012, Special Issue Section: Selected Papers from the 1st African-European Conference on Chemometrics, Rabat, Morocco, September 2010 Special Issue Section: Preprocessing methods Special Issue Section: Spectroscopic imaging, ISSN: 0169-7439. DOI: <https://doi.org/10.1016/j.chemolab.2012.02.002>. dirección: <https://www.sciencedirect.com/science/article/pii/S0169743912000251>.
- [7] Å. Rinnan, "Pre-processing in vibrational spectroscopy – when, why and how", *Analytical Methods*, vol. 6, pág. 7124, 2014. DOI: [10.1039/C3AY42270D](https://doi.org/10.1039/C3AY42270D).
- [8] C. E. Alciaturi, M. E. Escobar, C. De La Cruz y C. Rincón, "Partial least squares (PLS) regression and its application to coal analysis", es, *Revista tecnica de la Facultad de*



- Ingeniería Universidad del Zulia*, vol. 26, págs. 197-204, dic. de 2003, ISSN: 0254-0770. dirección: [http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S0254-077020030003000006&nrm=iso](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0254-077020030003000006&nrm=iso).
- [9] L. R., A. . y A. ., “Estudio comparativo de modelos de predicción obtenidos para propiedades de calidad en suelo, frutas y vegetales utilizando las técnicas de análisis multivariantes PLS y PCR”, Español, *Revista Ciencias Técnicas Agropecuarias*, vol. 18, págs. 50-54, 2009, ISSN: 1010-2760. dirección: <https://www.redalyc.org/articulo.oa?id=93215240011>.
- [10] J. C. Vega-Vilca y J. Guzmán, “Regresión PLS y PCA como solución al problema de multicolinealidad en regresión múltiple”, *Rev. Mat*, vol. 18, n.º 1, págs. 09-20, 2011, Electronic Document Format(ISO), ISSN: 1409-2433. dirección: [http://www.scielo.sa.cr/scielo.php?script=sci\\_arttext&pid=S1409-243320110001000002&lng=en&nrm=iso](http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1409-243320110001000002&lng=en&nrm=iso).
- [11] P. D. Wentzell y L. Vega Montoto, “Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures”, *Chemometrics and Intelligent Laboratory Systems*, vol. 65, n.º 2, págs. 257-279, 2003, ISSN: 0169-7439. DOI: [https://doi.org/10.1016/S0169-7439\(02\)00138-7](https://doi.org/10.1016/S0169-7439(02)00138-7). dirección: <https://www.sciencedirect.com/science/article/pii/S0169743902001387>.
- [12] Á. F. Egaña, F. A. Santibáñez-Leal, C. Vidal, G. Díaz, S. Liberman y A. Ehrenfeld, “A Robust Stochastic Approach to Mineral Hyperspectral Analysis for Geometallurgy”, *Minerals*, vol. 10, n.º 12, 2020, ISSN: 2075-163X. DOI: [10.3390/min10121139](https://doi.org/10.3390/min10121139). dirección: <https://www.mdpi.com/2075-163X/10/12/1139>.
- [13] C. Schropp, H. Kanpp y K. K. H. Neubert, “Potential of NIR hyperspectral imaging in the minerals industry”, en *OCM 2013 - Optical Characterization of Materials - conference proceedings / edited by Jürgen Beyerer / Fernando Puente León / Thomas Längle*, Karlsruhe: KIT Scientific Publishing, 2013, págs. 215-225. dirección: <https://publications.rwth-aachen.de/record/208866>.
- [14] C. Sandoval-Muñoz, G. Velásquez, J. Álvarez et al., “Enhanced elemental and mineralogical imaging of Cu-mineralized rocks by coupling  $\mu$ -LIBS and HSI”, *J. Anal. At. Spectrom.*, vol. 37, págs. 1981-1993, 10 2022. DOI: [10.1039/D2JA00147K](https://doi.org/10.1039/D2JA00147K). dirección: <http://dx.doi.org/10.1039/D2JA00147K>.
- [15] T. M. Inc., *MATLAB version: 9.6.0.1072779 (R2019a)*, Natick, Massachusetts, United States, 2019. dirección: <https://www.mathworks.com>.

- [16] N. Mobaraki y J. M. Amigo, “HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis”, *Chemometrics and Intelligent Laboratory Systems*, vol. 172, págs. 174-187, 2018, ISSN: 0169-7439. DOI: <https://doi.org/10.1016/j.chemolab.2017.11.003>. dirección: <https://www.sciencedirect.com/science/article/pii/S0169743917306305>.
- [17] V. Henríquez, “ANÁLISIS QUÍMICO CUANTITATIVO DE CONCENTRADOS DE COBRE UTILIZANDO MICROSCOPIA ÓPTICA MULTIESPECTRAL”, Universidad de Concepcion, Concepcion, 2020.
- [18] J. M. Amigo, “Chapter 1.1 - Hyperspectral and multispectral imaging: setting the scene”, en *Hyperspectral Imaging*, ép. Data Handling in Science and Technology, J. M. Amigo, ed., vol. 32, Elsevier, 2019, págs. 3-16. DOI: <https://doi.org/10.1016/B978-0-444-63977-6.00001-8>. dirección: <https://www.sciencedirect.com/science/article/pii/B9780444639776000018>.
- [19] I. Torres, J. Amigo Rubio, R. Ipsen, J. Amigo y R. Ipsen, “Using fractal image analysis to characterize microstructure of low-fat stirred yoghurt manufactured with microparticulated whey protein”, *Journal of Food Engineering*, vol. 109, págs. 721-729, 2012. DOI: [10.1016/j.jfoodeng.2011.11.016](https://doi.org/10.1016/j.jfoodeng.2011.11.016). dirección: <https://doi.org/10.1016/j.jfoodeng.2011.11.016>.
- [20] R. C. Gonzalez, R. E. Woods y S. L. Eddins, *Digital Image Processing Using Matlab*. 2004. dirección: <https://doi.org/10.1117/1.3115362>.
- [21] M. Vidal y J. Amigo, “Pre-processing of hyperspectral images. Essential steps before image analysis”, *Chemometrics and Intelligent Laboratory Systems*, vol. 117, págs. 138-148, 2012. DOI: [10.1016/j.chemolab.2012.05.009](https://doi.org/10.1016/j.chemolab.2012.05.009). dirección: <https://doi.org/10.1016/j.chemolab.2012.05.009>.
- [22] J. M. Amigo, I. Martí y A. Gowen, “Chapter 9 - Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality”, en *Chemometrics in Food Chemistry*, ép. Data Handling in Science and Technology, F. Marini, ed., vol. 28, Elsevier, 2013, págs. 343-370. DOI: <https://doi.org/10.1016/B978-0-444-59528-7.00009-0>. dirección: <https://www.sciencedirect.com/science/article/pii/B9780444595287000090>.

- [23] A. Gowen, C. O'Donnell, P. Cullen, G. Downey y J. Frias, "Hyperspectral imaging an emerging process analytical tool for food quality and safety control", *Trends in Food Science & Technology*, vol. 18, n.º 12, págs. 590-598, 2007, ISSN: 0924-2244. DOI: <https://doi.org/10.1016/j.tifs.2007.06.001>. dirección: <https://www.sciencedirect.com/science/article/pii/S0924224407002026>.
- [24] J. M. Amigo, "Practical issues of hyperspectral imaging analysis of solid dosage forms", *Analytical and Bioanalytical Chemistry*, vol. 398, n.º 1, págs. 93-109, 2010, ISSN: 1618-2650. DOI: [10.1007/s00216-010-3828-z](https://doi.org/10.1007/s00216-010-3828-z). dirección: <https://doi.org/10.1007/s00216-010-3828-z>.
- [25] P. Geladi, J. Burger y T. Lestander, "Hyperspectral imaging: calibration problems and solutions", *Chemometrics and Intelligent Laboratory Systems*, vol. 72, n.º 2, págs. 209-217, 2004, Advances in Chromatography and Electrophoresis - Conferentia Chemometrica 2003, Budapest, ISSN: 0169-7439. DOI: <https://doi.org/10.1016/j.chemolab.2004.01.023>. dirección: <https://www.sciencedirect.com/science/article/pii/S0169743904000371>.
- [26] B. Boldrini, W. Kessler, K. Rebner y R. W. Kessler, "Hyperspectral Imaging: A Review of Best Practice, Performance and Pitfalls for in-line and on-line Applications", *J. Near Infrared Spectrosc.*, vol. 20, n.º 5, págs. 483-508, 2012. dirección: <https://opg.optica.org/jnirs/abstract.cfm?URI=jnirs-20-5-483>.
- [27] J. M. Amigo, J. Cruz, M. Bautista, S. MasPOCH, J. Coello y M. Blanco, "Study of pharmaceutical samples by NIR chemical-image and multivariate analysis", *TrAC Trends in Analytical Chemistry*, vol. 27, n.º 8, págs. 696-713, 2008, ISSN: 0165-9936. DOI: <https://doi.org/10.1016/j.trac.2008.05.010>. dirección: <https://www.sciencedirect.com/science/article/pii/S0165993608001301>.
- [28] F. Firtha, A. Fekete, T. Kaszab et al., "Methods for Improving Image Quality and Reducing Data Load of NIR Hyperspectral Images", *Sensors*, vol. 8, n.º 5, págs. 3287-3298, 2008, ISSN: 1424-8220. DOI: [10.3390/s8053287](https://doi.org/10.3390/s8053287). dirección: <https://www.mdpi.com/1424-8220/8/5/3287>.
- [29] L. Zhang y M. J. Henson, "A Practical Algorithm to Remove Cosmic Spikes in Raman Imaging Data for Pharmaceutical Applications", *Applied Spectroscopy*, vol. 61, n.º 9, págs. 1015-1020, 2007, PMID: 17910800. DOI: [10.1366/000370207781745847](https://doi.org/10.1366/000370207781745847). eprint:

<https://doi.org/10.1366/000370207781745847>. dirección: <https://doi.org/10.1366/000370207781745847>.

- [30] J. M. Amigo y C. Santos, “Chapter 2.1 - Preprocessing of hyperspectral and multispectral images”, en *Hyperspectral Imaging*, ép. Data Handling in Science and Technology, J. M. Amigo, ed., vol. 32, Elsevier, 2019, págs. 37-53. doi: <https://doi.org/10.1016/B978-0-444-63977-6.00003-1>. dirección: <https://www.sciencedirect.com/science/article/pii/B9780444639776000031>.
- [31] I. Torres y J. M. Amigo, “Chapter 2.8 - An overview of regression methods in hyperspectral and multispectral imaging”, en *Hyperspectral Imaging*, ép. Data Handling in Science and Technology, J. M. Amigo, ed., vol. 32, Elsevier, 2019, págs. 205-230. doi: <https://doi.org/10.1016/B978-0-444-63977-6.00010-9>. dirección: <https://www.sciencedirect.com/science/article/pii/B9780444639776000109>.
- [32] C. Ravn, E. Skibsted y R. Bro, “Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—Comparing common calibration approaches”, *Journal of Pharmaceutical and Biomedical Analysis*, vol. 48, n.º 3, págs. 554-561, 2008, ISSN: 0731-7085. doi: <https://doi.org/10.1016/j.jpba.2008.07.019>. dirección: <https://www.sciencedirect.com/science/article/pii/S0731708508003944>.

## Anexo A

---

### Código para crear matriz X

---

```
h=1;
for P=1:140
    for Q=1:160
        for R=1:255
            Aux(h,R) = muestra{i}(P,Q,R);
        end
        h=h+1;
    end
end
Prom=nanmean(Aux);
aditionaltext='_prom';
newvariablename=[fileName , additionaltext];
assignin('base', newvariablename, Prom);
matrix=[Prom]; %Creacion Hiper-matriz
Xmatrix(i,:)=matrix;
```

---

### Código para crear matriz Y

---

```
[data , headers] = xlsread...
('Inventario_de_muestras_Septiembre_2020.xlsx', 4, '', 'Basic');
workspacevariables=who;
variableIndex = find(ismember(headers , workspacevariables));
desfase= 3;
Bornita           = data(variableIndex-desfase , 16);
Covelina          = data(variableIndex-desfase , 17);
Calcopirita       = data(variableIndex-desfase , 18);
Enargita          = data(variableIndex-desfase , 19);
Molibdenita       = data(variableIndex-desfase , 20);
Pirita            = data(variableIndex-desfase , 21);
Cuarzo            = data(variableIndex-desfase , 22);
OxidoDeHierro     = data(variableIndex-desfase , 23);
Otros             = data(variableIndex-desfase , 24);
```

---

**UNIVERSIDAD DE CONCEPCION – FACULTAD DE INGENIERIA**  
**RESUMEN DE MEMORIA DE TITULO**

**Departamento** : Departamento de Ingeniería Eléctrica  
**Carrera** : Ingeniería Civil Electrónica  
**Nombre del memorista** : Patricio Humberto Quezada Burgos  
**Título de la memoria** : Análisis de imágenes hiper espectrales en el rango infrarrojo de concentrado de cobre usando Hypertools  
**Fecha de la presentación oral** : 01/09/2023  
**Profesor(es) guía** : Daniel G. Sbarbaro H.  
**Profesor(es) revisor(es)** : Sebastián E. Godoy M., Luis E. Arias P.  
**Concepto** :  
**Calificación** :

**Resumen (máximo 200 palabras)**

La minería es una industria que ha crecido a pasos agigantados en el transcurso de los años y que se visto beneficiada por los grandes avances tecnológicos que se han ido desarrollando hasta la actualidad. Estas nuevas tecnologías han promovido el interés de la comunidad científica he impulsado una amplia variedad de estudios sobre la extracción de recursos mineros en donde se busca para aumentar su eficiencia y niveles de productividad.

En este estudio se aplicaron técnicas de regresión que permitieron mediante la extracción y manipulación de datos de imágenes hiper-espectrales de concentrados de cobre, obtener modelos que permitan explicar la variabilidad de la concentración porcentual de la mineralogía de ciertas muestras. Para lograr este objetivo se utilizó un conjunto de datos correspondientes a 32 imágenes hiper-espectrales tomadas a muestras de concentrados de cobre. A las que se le realizo un exhaustivo análisis, con el que se logró tener una estructura de datos muchos más limpia.

El trabajo se realizó en el software de Matlab en conjunto con la herramienta Hypertools, las cuales permitieron procesar los diversos datos de interés a estudiar, en ellos se implementó la aplicación del método de regresión por mínimos cuadrados (PLS) y nos entregó como resultado los modelos para cada uno de los minerales que se analizó.

Los resultados obtenidos dan cuenta de la capacidad adaptativa del modelo de regresión PLS, el cual generó modelos que lograron estimar con una precisión aceptable ( $R^2 > 0.70$ ) la variabilidad de las concentraciones de minerales entre las que se destacó las respuestas obtenidas en los minerales: Bornita, Calcopirita, Enargita, Piritita y Óxido de hierro, los que presentaron además los mejores resultados en términos de precisión al momento de enfrentarse a datos fuera del conjunto de entrenamiento.