

UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO INGENIERÍA DE MINAS

**MODELAMIENTO GEOMETALÚRGICO DE
RECUPERACIÓN DE COBRE INCORPORANDO DATOS
ESPECTRALES**

TESIS PARA OPTAR AL GRADO DE MAGISTER EN MINERÍA

GUSTAVO ANDRÉS DONOSO DROGUETT

PROFESOR GUÍA:

XAVIER EMERY

PROFESOR CO-GUÍA:

ALEJANDRO CÁCERES SAAVEDRA

MIEMBROS DE LA COMISIÓN:

NADIA MERY GUERRERO

MOHAMMADALI MALEKITEHRANI

SANTIAGO DE CHILE

2020

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE: MAGISTER DE MINERIA
POR: GUSTAVO DONOSO DROGUETT
FECHA: 30/06/2020
PROF. GUIA: XAVIER EMERY

MODELAMIENTO GEOMETALÚRGICO DE RECUPERACIÓN DE COBRE INCORPORANDO DATOS ESPECTRALES

La incertidumbre asociada a la ausencia de información en el ámbito minero depende de la etapa del proyecto y el tipo de información puede ser relevante para entender las variaciones mineralógicas, que rara vez son capturadas con suficiente nivel de detalle para entender y mitigar impactos negativos en la alimentación, recuperación y calidad del concentrado.

Esta tesis presenta una implementación del uso de información de espectrómetros para complementar un modelo de recuperación de cobre 3D.

Esto permitiría incorporar las variaciones mineralógicas en ensayos de recuperación de cobre a través de un entrenamiento supervisado, desafiando los enfoques tradicionales utilizados en geometalurgia.

La información que se expone para explicar estas variaciones mineralógicas se basó en mediciones de espectrómetros, generalmente, utilizados en las primeras etapas de exploración minera para la identificación de minerales de difícil identificación macroscópica.

El estudio describe una propuesta de integración de información de espectrómetros a los ensayos de flotación para interpretar la recuperación de cobre desde un punto de vista de mineralogía.

Este análisis entrega una metodología para integrar la información espectral en el modelo de bloques y una técnica de regresión para utilizar esta información de una manera alternativa.

AGRADECIMIENTOS

Agradezco el apoyo del Departamento de Ingeniería de Minas, al Advanced Mining Technology Center (AMTC) de la Universidad de Chile, así como la Agencia Nacional de Investigación y Desarrollo, a través del proyecto CONICYT / FONDECYT / REGULAR / 1170101.

TABLA DE CONTENIDO

1	INTRODUCCIÓN	1
1.1	OBJETIVOS	2
1.1.1	Objetivos Generales	2
1.1.2	Objetivos Específicos.....	2
1.2	ALCANCES	2
2	ESTADO DEL ARTE.....	4
3	ANTECEDENTES TEÓRICOS	7
3.1	Regresión Lineal	7
3.1.1	Regresión de Componentes Principales (PCR).....	7
3.1.2	Regresión de mínimos cuadrados parciales (PLSR)	12
3.2	Técnicas de cuantificación de minerales	14
3.2.1	Near InfraRed (NIR)	14
3.2.2	Quantitative Evaluation of Materials by Scanning (QEMSCAN)	16
3.2.3	X-Ray Diffraction (XRD)	16
3.3	Geoestadística	17
3.3.1	Definición de conceptos fundamentales.....	17
3.3.2	Variograma	19
3.3.3	Kriging	20
3.3.4	Kriging Ordinario.....	20
3.3.5	Cokriging Ordinario.....	21
3.4	Concentración de minerales	22
3.4.1	Concentración por Flotación de minerales.....	22
3.4.2	Colectores	23
3.4.3	Espumantes	23
3.4.4	Reguladores.....	24
3.4.5	Ensayos de flotación	25
4	METODOLOGÍA PROPUESTA	27
4.1	Etapa 1.- Regresión Multivariable	27
4.2	Etapa 2.- Predicción de un modelo 3D recuperación de cobre	28

5	CASO DE ESTUDIO	30
5.1	Etapa 1	30
5.1.1	Características generales de la base de datos integrada	30
5.1.2	Revisión de datos anómalos.....	30
5.1.3	Calibración.....	35
5.1.4	Validación del modelo regresión propuesto.....	39
5.2	Etapa 2 - Predicción de un modelo 3D de recuperación de cobre.....	41
5.2.1	Revisión de parámetros de cokriging.....	42
5.2.2	Revisión de la estimación de recuperación de cobre y componentes principales	43
5.2.3	Validación cruzada.....	45
5.2.4	Revisión del número óptimo de componentes	48
6	DISCUSIONES GENERALES	50
7	CONCLUSIONES	53
	BIBLIOGRAFÍA	55

INDICE DE ILUSTRACIONES

Figura 1 Las tres técnicas para medir señales de vibraciones; el eje de coordenada es Energía y el eje abscisas es longitud de onda (tomado de (Donald Burns; Emil Ciurczak, 2008)).	14
Figura 2 Vibraciones armónicas y vibraciones combinadas (derecha), una representación matemática de RTT (Radiative Transfer Theory) tomado de (Bioucas-Dias José et al, 2012).	15
Figura 3 Ejemplo de Reflectancia espectral para cuatro minerales de hierro. tomado de (Bioucas-Dias José et al, 2012).	15
Figura 4 Ejemplo de función de distribución acumulada (cdf). Fuente Apunte de Geoestadística - Emery.	18
Figura 5 Ejemplo de variograma experimental en direcciones horizontal y vertical. FUENTE: (MI68A -Emery).	19
Figura 6 Clasificación de colectores ionizantes de acuerdo al tipo anión, catión y sus aplicación (tomado de (Wills, 2006)).	24
Figura 7 Máquina de flotación de laboratorio Denver Metso D-12.	26
Figura 8 Esquema conceptual de integración de información de espectrómetros y datos de flotación de cobre.	28
Figura 9 Esquema conceptual de integración de espectros al modelo 3D propuesta para predecir la recuperación de Cobre.	29
Figura 10 Muestra datos de entrenamiento en (a) %Cu en alimentación; (b) masa de alimentación.	31
Figura 11 Muestra las características de %RecCu utilizado para la calibración del modelo: (a) histograma de %RecCu (b) gráfico de probabilidad normal de %RecCu.	32
Figura 12 Distancias de Mahalanobis con 10 componentes. (a) PLSR obtenidas desde los datos de entrenamiento, (b) PCA.	33
Figura 13 Distancia de Cook para el conjunto de entrenamiento.	34
Figura 14 Muestra outliers sugeridos usando distancia de Cook (a) registro #308 (MPHYP-589) con 58%RecCu y (b) registro #101 (MPHYP-245) con 93%RecCu.	34
Figura 15 Cuadro comparativo de QEMSCAN para ambos registros con bajo desempeño MPHYP-245 y MPHYP-589.	35
Figura 16 Comparación de métodos de regresión de datos de calibración de recuperación de cobre: (a) porcentaje de varianza explicado según número de componentes, (b) nube de correlación entre recuperación de cobre real y estimada.	36
Figura 17 Scatter plot de los %RecCu reales versus los estimados usando (a) PLSR y (b) PCA.	36
Figura 18 Errores residuales entre valores reales y valores estimados. En (a) para PLSR y en (b) para PCA.	36
Figura 19 Muestra reflectancia NIR para los registros #179(MPHYP-356.asd); #271 (MPHYP-496.asd); #180(MPHYP-357.asd).	38
Figura 20 Muestra la asignación de los ponderadores para la generación de variables.	39
Figura 21 (a) Gráfico de dispersión del %RecCu usando PLSR. (b) Errores residuales las predicciones propuestas.	39

Figura 22 (a) Gráfico de dispersión del %RecCu usando PCR. (b) Errores residuales las predicciones propuestas.	40
Figura 23 Vista en planta de la información del depósito. (a) Información espectral disponible. (b) Zona de trabajo para el modelo 3D.	41
Figura 24 Muestra en (a) el variograma experimental directo de %RecCu y (b) el variograma experimental cruzado entre %RecCu y la componente principal 1.	42
Figura 25 Variograma modelados (a) modelamiento de componente principal (CP) uno y tres; (b) modelamiento de vario grama de CP1 y Recuperación de cobre.....	43
Figura 26 Muestra un histograma de %RecCu de las estrategias de estimación. Una manera directa y otra manera indirecta.....	44
Figura 27 Nube de dispersión entre los métodos de estimación directa e indirecta.....	44
Figura 28 Gráficos de probabilidad normal de los resultados de validación cruzada. Valore real ensayado (azul), método directo (rojo), y método indirecto (verde).....	46
Figura 29 Muestra gráficos de dispersión de la estimación de %RecCu. (a) Método directo; (b) método indirecto	46
Figura 30 Errores residuales obtenidos para la validación cruzada	47
Figura 31 Información de especies minerales mayoritarias en espectros con mal desempeño....	48
Figura 32 (a) Box-plot de errores asociados a la validación cruzada en función del número de componentes. (b) Gráfico de media según el número de componentes.	49
Figura 33 Muestra los resultados de la validación cruzada entre espectro medido y estimación de cokriging	51
Figura 34 Muestra dos minerales de ganga Biotita(BioMA1) e Ilita(IIAZ2f) de librería Specmin@Pro.....	52

INDICE DE TABLAS

Tabla 1 Esquema de reactivos para ensayos de flotación	30
Tabla 2 Estadísticas básicas de las leyes de cobre de los ensayos de flotación <i>rougher</i>	31
Tabla 3 Estadísticas básicas de masa de los ensayos de flotación <i>rougher</i>	31
Tabla 4 Estadísticas básicas de resultados de %RecCu estimada	44
Tabla 5 Estadísticas básicas de los resultados de la validación cruzada	45
Tabla 6 Estadísticas básicas de errores asociados a la validación cruzada en función de los 176 ensayos de flotación	49

1 INTRODUCCIÓN

La incertidumbre asociada a la ausencia de información en el ámbito minero depende de la etapa del proyecto y, por otro lado, la adquisición de información consume tiempo, utiliza recursos económicos y además, requiere procesamiento. Este estudio presenta una implementación del uso de información de espectrómetros para complementar un modelo de recuperación de cobre 3D basado en un entrenamiento supervisado.

La solución a la ausencia de información es la utilización de variables “proxy”, es decir, datos complementarios que, ciertamente, no tienen gran interés, pero que sirven para obtener una base informativa de mucho mayor interés.

La propuesta de este estudio, fue la utilización de una información complementaria a los ensayos de flotación como son los datos de espectrometría.

Los espectrómetros son instrumentos ópticos que miden propiedades de luz sobre un elemento. Los minerales se relacionan de manera distinta frente a la luz, ya sea absorbiéndola o reflejándola. Permiten la identificación de estos y, por otro lado, los resultados de ensayos de concentración por flotación de cobre se ven alterados con la presencia de ciertos minerales que perjudican su desempeño.

Los ensayos de flotación por batch o lote se realizan en condiciones ideales de cantidad de masa, granulometría de la muestra, contenido de reactivos y volumen de agua.

La caracterización de la masa utilizada en el ensayo tiene información que permitirá explicar el desempeño de las pruebas en términos de recuperación del mineral de interés. La literatura asociada a la cinética de los ensayos de flotación de cobre tiene bien documentado el impacto de los minerales de ganga en el desempeño de las pruebas que se han desarrollado.

El estudio describe una propuesta de integración de información de espectrómetros a los ensayos de flotación para interpretar la recuperación de cobre desde un enfoque de mineralogía de ganga, pero ahora en un contexto espacial de un modelo 3D.

1.1 OBJETIVOS

1.1.1 Objetivos Generales

- Implementar un modelo de estimación 3D con el porcentaje de recuperación de cobre (%RecCu) a partir de datos de espectrómetros y ensayos de flotación.

1.1.2 Objetivos Específicos

- Evaluar alternativas de algoritmos de regresión para integrar mediciones de espectrómetros y resultados de ensayos de flotación *rougher*.
- Evaluar alternativas de predicción de recuperación de cobre como un modelo 3D usando las mediciones disponibles de espectrómetros y resultados de ensayos de flotación *rougher*.

1.2 ALCANCES

El alcance del presente análisis y estudio responde a una necesidad de la industria minera para incorporar información adicional de bajo costo que permita mejorar su desempeño operacional.

El problema planteado se puede resumir en una necesidad de utilizar información capturada por instrumentos ópticos, comúnmente conocido como espectrómetros, para mejorar la predicción de recuperación de cobre en zonas sin información de ensayos de flotación.

Los espectrómetros son instrumentos ópticos que usan detectores para medir la distribución de radiación en una región particular de longitud de onda. Los datos medidos por un espectrómetro se denominan espectros. En esta oportunidad se utilizó un dispositivo denominado LabSpec distribuido por la compañía ASD Inc.

Este aparato es capaz de medir la radiación en dos regiones; infrarojo cercano visible (VNIR:350-1000nm) e infrarojo cercano (NIR:1001-2500nm).

Los datos espectrales provienen de una mina de Cu-Mo de una operación ubicada en la región de Antofagasta - Chile. Los registros de espectrómetros provienen de sondajes, pozos de tronadura y muestreo sistemático desde la correa transportadora y fueron medidos con el dispositivo LabSpec a una granulometría estandarizada de 10# (1.651 mm).

Los espectros fueron trabajados como reflectancia relativa, es decir, se calcula como la razón entre la luz reflectada sobre una muestra y un blanco de referencia, siendo este último un material sin capacidad de absorber luz o 100% reflectante a la luz recibida.

Las herramientas de regresión que se utilizaron son regresión de componentes principales (PCR) y regresión de mínimo cuadrados (PLSR, de las siglas en inglés *Partial Least Squared Regression*).

La herramienta utilizada para desarrollar el trabajo geoestadístico es una librería implementada en el software Matlab.

Así mismo, para desarrollar el trabajo lectura de archivos espectrales, se utilizó una librería llamada *Field Spectroscopy Facilities Post Processing* implementada en el software Matlab (Robinson, Iain; MacArthur, Chris, 2015).

La interpretación de espectros fue realizada con el software Specmin® Pro versión 3.1.

El software contiene una base de datos con más de 1.500 espectros y se puede destacar que tiene una librería de espectros para pórfidos de cobre.

2 ESTADO DEL ARTE

La geo metalurgia es una disciplina que integra geología, metalurgia y experiencia operacional. Sin embargo, las variaciones mineralógicas pueden no ser capturadas con suficiente detalle para entender o mitigar impactos negativos en alimentación, recuperación y calidad de concentrado.

La solución implementada en geo metalurgia, es el etiquetado de los materiales según su desempeño en las plantas de procesamiento. La calidad de roca o calidad de los materiales es una solución ampliamente usada en la industria minera.

Los enfoques más utilizados son el modelamiento de abundancia de minerales o el modelamiento de variables categóricas y la complejidad de los algoritmos utilizados es variada debido al gran desarrollo de nuevas técnicas de modelamiento.

Las distribuciones de abundancia de minerales pueden ser estimadas utilizando variables cualitativas generadas del mapeo de pozos de tronadura (Cornejo, 2009). Para desarrollar este trabajo se modeló y co-simuló las variables correlacionadas como indicadores de la presencia de minerales de mena.

Estas eran variables continuas truncadas a un número limitado de clases (Emery & Cornejo, 2010). Por otro lado, las limitaciones observadas son la aplicación de intervalos discretos para caracterizar porcentajes en indicadores de intensidad, si bien esto permite reducir la variabilidad de los datos muestreados, lo hace sin considerar la forma de la distribución de los datos lo que genera intervalos con baja frecuencia de datos en desmedro de intervalos de alta frecuencia.

Asimismo, algunos trabajos fueron desarrollados para ligar el contenido de una especie mineralógica al contenido de cobre total con las variables categóricas (litología). Por ejemplo, la reproducción de incertidumbre bajo un modelo híbrido de variables categóricas y variables continuas correlacionadas espacialmente.

Para las variables continuas fueron representadas a través de una transformación Gaussiana, mientras que la variable categórica fue desarrollada como una Gaussiana truncada (Emery & Silva, 2009). Algunos trabajos en donde se compararon distintos modelos en los que se correlacionaron estos datos muestran como resultado una reducción de la incertidumbre en el contenido de cobre, cuando se utiliza la información geológica para co-simular (Caceres, 2010). Debido a que la variabilidad de las leyes contenidas en algunas litologías es significativa con respecto a la población global de los datos, la proporción de una variable categórica incide en una menor dispersión de valores en contenido de cobre total.

Relativo a la proporción y a las relaciones de contacto, una interpretación de las relaciones de contacto entre las unidades litológicas para el depósito de Mansa Mina, el trabajo aplicó los métodos de simulación secuencial de indicadores y simulación plurigaussiana (De la Quintana, 2008). En la misma línea, la utilización de modelos determinísticos 3D en la industria minera es una actividad habitual, un trabajo de simulación de variables categóricas comparado con un modelo determinístico (wireframe 3D), para ello se usaron dos algoritmos distintos: una simulación multipunto llamada “*single normal equation*” (SNESIM) y una simulación secuencial de indicadores (BlockSIS), donde SNESIM muestra una mejor reproducción de los límites de contacto (Hurtado, 2009).

La aplicación de herramientas geoestadísticas también ha enfocado sus esfuerzos a la estimación de variables categóricas en el área de mecánica de rocas (Egaña, 2008).

En particular, se enfrenta a la estimación de variables no aditivas y es el caso la estimación del índice de RMR (rock mass rating), simulando cada uno de sus componentes de manera independiente como un indicador de incertidumbre. Las limitaciones de la metodología propuesta es la dificultad de correlacionar las variables categóricas entre sí, es decir, una roca con alto índice de fracturamiento, también estará ligada a un índice de dureza bajo. Sin embargo, al ser simulados de manera independiente se puede llegar a un mismo puntaje (rating) pero con distinta combinación de factores.

En los últimos años ha cobrado fuerza la utilización de herramientas que permitan el procesamiento de un gran volumen de información, la industria minera no se ha quedado atrás, si bien existen algunos trabajos enmarcados bajo el término de “Minería de datos”.

Por “Minería de datos” (Data Mining) se debe entender como la extracción implícita, previamente desconocida y potencialmente útil a partir de los datos (Witten & Eibe, 2005). Básicamente es un proceso analítico diseñado para explorar una gran cantidad de información en la búsqueda de patrones y/o relaciones sistemáticas entre las variables.

Dentro de los trabajos con aplicación para la industria minera en el área de procesos, se tiene un modelo predictivo de recuperaciones de cobre, molibdeno y leyes de concentrado colectivo mediante técnicas de minería de datos como redes neuronales (Dirección de Modelamiento Geometalúrgico – Codelco Norte, 2011), ha sido capaz de predecir la recuperación real diaria en planta concentradora utilizando variables predictoras geológicas como mineralogías, leyes y alteraciones. También se pueden mencionar un trabajo desarrollado en minería de datos, para el mejoramiento de la interpretación geológica mediante el uso de técnicas estadísticas multivariadas (Carmona, 2009).

La industria minera utiliza instrumentación óptica en las etapas de exploración minera como una herramienta para distinguir minerales que macroscópicamente suelen ser imposibles de diferenciar. El grupo de los filosilicatos suelen ser de difícil identificación por los distintos tipos de sustitución de Al, Mg, Fe de su estructura tetraédrica. Una herramienta muy útil para su diferenciación ha sido el uso de espectrómetro. Más detalles pueden ser encontrado en libros especializados (Velde, B. & Meunier, Alain, 2008), (Velde, 1977).

Phoenix mine publicó un trabajo de integración de datos de espectrómetros en pozos de tronadura y muestreos en planta para un depósito de Cu-Au. Su objetivo fue complementar la información operacional de recuperación y capacidad de molienda de los equipos.

Un modelo de recuperación y alimentación planta (capacidad de molienda) fue implementado y su validación fue realizada con datos operacionales. El trabajo propuso tres métodos de regresión, aunque no se trató de un modelo 3D. Sus autores (Johnson, Curtis & Browning, David & Pendock, Neil, 2019) restringieron a 41 intervalos del total de la longitud de onda.

Un estudio investigó el efecto de caolinita con baja y alta cristalinidad. Porque Los minerales arcillosos modifican la reología de la pulpa debido al tipo de estructura cristalográfica, estas al hidratarse aumenta significativamente la viscosidad generando impacto negativo en la recuperación en minerales de cobre y oro. (Zhang & Peng, 2015)

Efectos de la bentonita y el caolinita en la recuperación y ley del concentrado de cobre en la flotación del mineral fueron bien documentado con rangos desde un 75% a 25% de variación en el porcentaje de recuperación de Cobre. (Y. Wang, Y. Peng, T. Nicholson, and R. Andreas, 2015)

Los efectos del tamaño de partículas para una concentración fija de arcillas fueron revisados para 190 [µm] con rangos entre 160 [µm] and 230 [µm] la recuperación puede tener variaciones de 10% del porcentaje de recuperación de cobre. En la medida que se reduce el tamaño de partícula mejora la recuperación. (Vilches, 2018)

La presencia de caolinita produce polarización y despolarización de la calcopirita, ya sea en agua destilada, agua de mar y agua destilada en presencia de los principales iones presentes en agua de mar. (Ortiz, 2017)

3 ANTECEDENTES TEÓRICOS

3.1 Regresión Lineal

3.1.1 Regresión de Componentes Principales (PCR)

El Análisis de Componentes Principales (PCA) fue introducido por Karl Pearson en 1901 y formulado por Hotelling en 1933. Consiste un método no paramétrico para extraer información relevante desde conjuntos de datos.

PCA proporciona una metodología para reducir un conjunto de datos complejos a una dimensión inferior. La metodología puede ser entendida como una transformación de las variables originales en otras nuevas y que no están correlacionadas entre ellas (las nuevas variables).

Sean las matrices X e Y de tamaño $m \times n$, y se relacionan a través de una transformación lineal P.

$$PX = Y$$

Donde:

X: son variables independientes (siempre disponible), x_i son columnas de X de $m \times 1$

P: matriz de transformación, p_i son filas P de $1 \times n$

Y: son variables dependientes (para ser predicho más tarde de X), y_i son columnas de Y

$$PX = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} [x_1 \quad \dots \quad x_n]$$

$$PX = \begin{bmatrix} p_1 x_1 & \dots & p_1 x_n \\ \vdots & \ddots & \vdots \\ p_m x_1 & \dots & p_m x_n \end{bmatrix}$$

Cada fila y_i es el resultado del producto punto entre columnas de X con la correspondiente fila en P.

$$y_i = \begin{bmatrix} p_1 \cdot x_i \\ \vdots \\ p_m \cdot x_i \end{bmatrix}$$

El ruido y la redundancia están presentes en el sistema lineal Y. El ruido usualmente tiene baja varianza y no tiene información relevante. Por otro lado, la redundancia tiene de alta varianza y permite realizar un cambio de base más concisa, lo que permite la reducción de dimensión.

El cálculo de la varianza para dos conjuntos de datos A y B con media cero, se definen de manera individual como el producto punto entre estas mismas.

$$A = \{a_1, a_2, \dots, a_n\}, B = \{b_1, b_2, \dots, b_n\}$$

$$\sigma_A^2 = \langle a_i a_i \rangle_i, \quad \sigma_B^2 = \langle b_i b_i \rangle_i$$

La covarianza entre dos variables A y B se define como el producto punto entre estas (variables). Hay dos aspectos importantes, si la covarianza es cero, A y B no están correlacionados. De manera equivalente, la covarianza de A y B puede ser formulado como un producto punto de matrices.

$$\sigma_{ab}^2 = \frac{1}{n-1} ab^T$$

Sea X una matriz de $m \times n$, y para cada fila de X corresponde x_i . La definición de la matriz de covarianza se escribe de la siguiente forma.

$$S_x = \frac{1}{n-1} XX^T$$

S_x es una matriz cuadrada simétrica de $m \times m$ y los términos de la diagonal de S_x son la varianza de x_i , por otro lado, todos los valores fuera de la diagonal de S_x corresponde a la covarianza entre los tipos de mediciones de X_i , es decir, las combinaciones de pares ordenados X.

La utilización de la metodología de PCA requiere del entendimiento los siguientes supuestos:

- La estructura propuesta por PCA supone una **linealidad** para Y (variables dependientes), es decir, Y puede ser escrito como una combinación lineal. Sin embargo, los sistemas complejos son casi siempre no lineales, y a menudo sus principales características cualitativas son el resultado directo de su no linealidad, aunque la condición de linealidad puede ser considerada una buena aproximación.
- La estructura propuesta por PCA supone una **jerarquización de las varianzas**, porque asume que las direcciones con mayor varianza son las más importante y están asociadas redundancia, mientras que componentes principales con baja varianza serán asociadas a ruido.
- La estructura propuesta por PCA supone una ortogonalidad de los vectores, las componentes principales son **ortogonales**, es decir, sus componentes no pueden ser una combinación lineal de otras, concepto que denominaremos linealmente independiente.

Las soluciones algebraicas de PCA están basadas en las propiedades de descomposición de vectores propios y/o descomposición de valores singulares.

3.1.1.1 Solución usando vectores propios de covarianza

Sea X una matriz de un conjunto de datos de tamaño $m \times n$, siendo m el número de mediciones y n el número de ensayos. El objetivo consiste en resumir la información Y a través de una matriz P que denominaremos de transformación.

$$Y = PX$$

Donde:

X son variables independientes (siempre disponible), x_i son columnas de X

P : matriz de transformación, p_i son filas P

Y son variables dependientes (para ser predicho más tarde de X), y_i son columnas de Y

$$S_Y = \frac{1}{n-1} YY^T$$

Donde:

S_Y = matriz de covarianza

n = número de ensayos usado para normalización

La matriz de covarianza puede ser reescrita en términos de una matriz de transformación P , que geoméricamente representa una rotación y un estiramiento o contracción de la magnitud del vector X .

$$S_Y = \frac{1}{n-1} YY^T$$

$$S_Y = \frac{1}{n-1} (PX)(PX)^T$$

$$S_Y = \frac{1}{n-1} PXX^T P^T$$

$$S_Y = \frac{1}{n-1} P(XX^T)P^T$$

$$S_Y = \frac{1}{n-1} PAP^T$$

Se define la matriz $A \equiv XX^T$ como una matriz simétrica y toda matriz cuadrada simétrica con coeficientes reales es ortogonalmente diagonalizable.

$$A = EDE^T$$

Donde:

D = una matriz diagonal

E = una matriz de vectores propios de A ordenados como columnas.

A = una matriz ortogonal de sus vectores propios $r \times m$.

Sea A una matriz con $r < m$, donde r es el rango de la matriz con vectores ortogonales y siendo m el número de variables. Dado que toda matriz cuadrada simétrica con coeficientes reales es ortogonalmente diagonalizable, se debe completar esta matriz con ceros, asignando cero a estos vectores $(m-r)$ para completar la matriz de $m \times m$.

La matriz \mathbf{P} es una matriz donde cada fila \mathbf{p}_i es un vector propio de $\mathbf{X}\mathbf{X}^T$. sustituyendo $\mathbf{P} \equiv \mathbf{E}^T$, se reescribe $\mathbf{A} = \mathbf{P}^T \mathbf{D} \mathbf{P}$. La inversa de una matriz ortonormal es su transposición.

$$S_Y = \frac{1}{n-1} \mathbf{P} \mathbf{A} \mathbf{P}^T$$

$$S_Y = \frac{1}{n-1} \mathbf{P} (\mathbf{P}^T \mathbf{D} \mathbf{P}) \mathbf{P}^T$$

$$S_Y = \frac{1}{n-1} (\mathbf{P} \mathbf{P}^T) \mathbf{D} (\mathbf{P} \mathbf{P}^T)$$

$$S_Y = \frac{1}{n-1} (\mathbf{P} \mathbf{P}^{-1}) \mathbf{D} (\mathbf{P} \mathbf{P}^{-1})$$

$$S_Y = \frac{1}{n-1} \mathbf{D}$$

Los resultados de PCA son las matrices \mathbf{P} y \mathbf{S}_Y . Las componentes principales de \mathbf{X} son los valores propios de $\mathbf{X}\mathbf{X}^T$ o las filas de \mathbf{P} .

Algoritmo A.1: PCA usando Covarianza

```
% función [signals,PC,V] = pca1(X)
% X      - matriz de mxn de datos de entrada
% signals - matriz mxn de datos proyectados
% PC     - cada columna componente principal
% V      - matriz mx1 de varianzas
Calcule cov = 1/(n - 1) X X^T
Calcule Obtenga [PC, V] = eig(cov)
% extraer diagonal de la matriz como un vector
Extraiga V = diag(V)
% ordenen las varianzas en orden decreciente
Obtenga [junk, rindices] = sort(-1*V);
Obtenga V=V(rindices)
Obtenga PC=PC(rindices)
% proyectar los datos originales
```

Determine $signals = PC * X^T$.
Finalice.

3.1.1.2 Solución usando Descomposición de Valores Singulares (SVD - Singular Value Decomposition)

Una solución algebraica de PCA consiste en la descomposición de matrices usando SVD.

Sea X una matriz cualquiera $n \times m$ y la matriz $X X^T$ de rango r , simétrica cuadrada de $n \times n$.

- $\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_r\}$ es un conjunto de vectores ortonormales de vectores propios de tamaño $m \times 1$, los valores propios asociados de la matriz simétrica $X X^T$ son $\{\widehat{\lambda}_1, \widehat{\lambda}_2, \dots, \widehat{\lambda}_r\}$.

$$(X X^T)\hat{v}_i = \lambda_i \hat{v}_i$$

- $\sigma_i \equiv \sqrt{\lambda_i}$ es un real positivo y denominado valores singulares.
- $\{\hat{u}_1, \hat{u}_2, \dots, \hat{u}_r\}$ es un conjunto de vectores ortonormales de $n \times 1$ definidos como $\hat{u}_i \equiv \frac{1}{\sigma_i} X \hat{v}_i$

Sea una matriz cualquiera X de tamaño $m \times n$, la matriz simétrica $X^T X$ tiene un conjunto de vectores propios ortonormales $\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_r\}$ y un conjunto de valores propios $\{\widehat{\lambda}_1, \widehat{\lambda}_2, \dots, \widehat{\lambda}_r\}$. El conjunto de vectores $\{X \hat{v}_1, X \hat{v}_2, \dots, X \hat{v}_r\}$ entonces, forma una base ortogonal, donde cada vector $X \hat{v}_i$ es de largo $\sqrt{\lambda_i}$.

- $\hat{u}_i \hat{u}_j = \delta_{ij}$
- $\|X \hat{v}_i\| = \sigma_i$

Las propiedades anteriores provienen del teorema de algebra lineal. Por lo tanto, re escribiendo de la siguiente forma.

$$X \hat{v}_i = \sigma_i \hat{u}_i$$

Donde:

X = matriz de $m \times n$

\hat{v}_i = vectores ortonormales de tamaño de $n \times r$

σ_i = valores propios matriz $r \times r$

\hat{u}_i = vectores propios de tamaño $m \times r$

La cantidad de vectores propios es $r \times m$ mientras que la dimensión de X es $n \times m$, siendo la matriz diagonal de $\{\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_r\}$ que se denominara Σ .

$$XV = U \Sigma$$

$$XVV^T = U \Sigma V^T$$

$$X = U \Sigma V^T$$

Algoritmo A.2: PCA usando SVD

```

% function [signals,PC,M] = pca2(X)
% X      matriz de mxn de datos de entrada
% signals - matriz mxn de datos proyectados
% PC     - cada columna componente principal
% M      - matriz mx1 de varianzas

% construir la matriz Y
Calcule  $Y = X^T / (\text{sqrt}(n - 1))$ 
% calcular la descomposición de matriz  $Y = USV^T$ 
Calcule  $[U,S,V] = \text{SVD}(Y)$ 
% extraer diagonal de la matriz como un vector
Extraiga  $S = \text{diag}(V)$ 
% ordenen las varianzas en orden decreciente
Obtenga  $M=S.*S$ 
Obtenga  $PC=V$ 
% proyectar los datos original
Determine  $\text{signals} = PC^T * X$ .
Finalice.

```

Lecturas recomendadas para profundizar estos temas: (Hubert, 2002) (Smith, 2002) (Shlens, 2003).

3.1.2 Regresión de mínimos cuadrados parciales (PLSR)

Regresión de mínimo cuadrado parciales (PLSR, del inglés Partial Least Squared Regression) intenta capturar la máxima varianza de X y alcanzar la máxima correlación entre X e Y .

$$X = TP^T + E$$

Donde:

X = matriz de variables independientes $n \times m$

P = matriz de X puntajes "scores" $m \times a$

T = matriz de X pesos "loadings" $n \times a$

E = error residual de X ($n \times m$)

Se puede construir la relación externa para el bloque Y de la misma manera como:

$$Y = UQ^T + F$$

Donde:

Y = matriz de variables dependientes $n \times p$

Q = matriz de Y puntajes "scores" $p \times a$

U = matriz de Y pesos "loadings" $n \times a$.

F = error residual de Y ($n \times p$)

La relación interna entre los puntajes del bloque Y , y los puntajes de bloque X a través de la relación de los vectores u y t .

$$\hat{u}_h = b_h t_h$$

Donde:

\hat{u}_h = una columna de vectores de puntaje “scores” del bloque Y , factor subíndice h ($n \times 1$)

b_h = los coeficientes de la regresión para una componente de PLS

t_h = una columna de vectores de puntaje “scores” del bloque X , factor subíndice h ($n \times 1$)

NIPALS (del inglés, Non Linear Iterative Partial Least Squares) es un algoritmo estándar para el cálculo de los componentes de PLSR y fue introducido por Wold (Wold, 1975).

Algoritmo A.3: buscando los k mayores vectores de componentes principales

Para $h = 1, \dots, c$, donde $A_0 = X^T Y$, $M_0 = X^T X$, $C_0 = I$, y c dado:

Calcule q_h , el vector propio dominante de $A_h^T A_h$.

Obtenga $w_h^* = C_h A_h q_h$.

Determine $w_h = w_h^* / \|w_h^*\|$.

Establezca $p_h^* = M_h w_h$.

Calcule $c_h = w_h^T M_h w_h$.

Obtenga $p_h = p_h^* / c_h$.

Determine $q_h = A_h^T w_h / c_h$.

Almacene w_h en W (como columna).

Guarde p_h en P (como columna).

Ponga q_h en Q (como columna).

Establezca $A_{h+1} = A_h - c_h p_h q_h^T$.

Calcule $M_{h+1} = M_h - c_h p_h p_h^T$.

Determine $C_{h+1} = C_h - w_h p_h^T$.

Finalice.

Lecturas recomendadas para profundizar estos temas: (Wold, 1975) (Geladi & Kowalski, 1986)

3.2 Técnicas de cuantificación de minerales

3.2.1 Near InfraRed (NIR)

La espectrometría de reflectancia del infrarrojo cercano (NIR) se está utilizando cada vez más como técnica de laboratorio para suplir análisis de XRD, basada en los conceptos de absorción y reflectancia de luz características propias de cada material.

El principio básico de espectroscopia es la vibración de los compuestos químicos y, aunque existen tres técnicas para medir la señal muy distintas, el principio es el mismo. Las técnicas que miden señal son MIR, NIR y Raman, siendo las dos primeras basadas en la absorción de una radiación.

El fenómeno de absorción considera un modelo de oscilador armónico donde dos masas vibratorias m_1 y m_2 (Figura 1) conducen a cambios de la distancia internuclear $<10\%$. Las absorciones de NIR se basan en vibraciones armónicas y vibraciones combinadas de acción de la molécula investigada. (Donald Burns; Emil Ciurczak, 2008).

Cuando un fotón de una fuente de luz incide en un material, la primera reflexión de superficie resultante puede ser descrita por la física clásica a través de las ecuaciones de Maxwell. Sin embargo, parte de la luz ingresa al material y su destino es sub-Objeto a las reglas de la mecánica cuántica, que determinan cómo reaccionará el material a la luz de varias longitudes de onda (Goetz, A.F.H., et al, 2008).

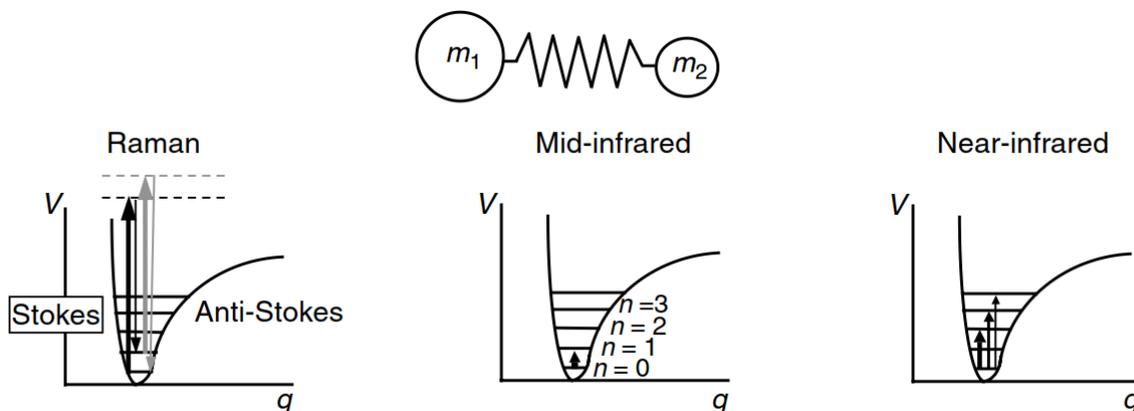


Figura 1 Las tres técnicas para medir señales de vibraciones; el eje de coordenada es Energía y el eje abscisas es longitud de onda (tomado de (Donald Burns; Emil Ciurczak, 2008)).

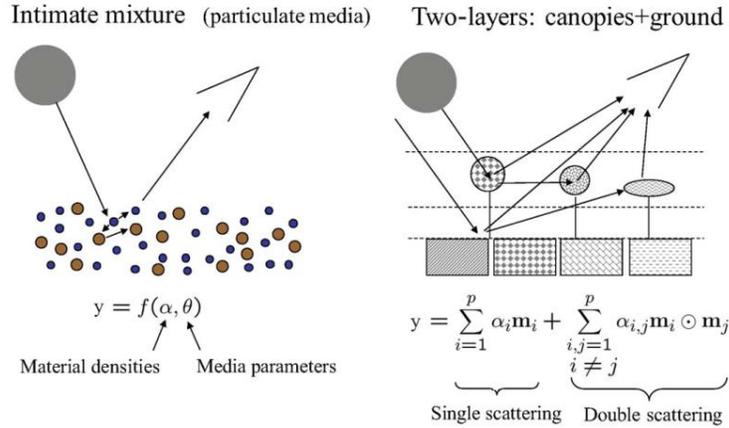


Figura 2 Vibraciones armónicas y vibraciones combinadas (derecha), una representación matemática de RTT (Radiative Transfer Theory) tomado de (Bioucas-Dias José et al, 2012)

El funcionamiento de los instrumentos de espectroscopía NIR implica la medición de un haz incidente reflejado desde la superficie de la muestra en varias longitudes de onda y desde una superficie reflectora estándar en las mismas longitudes de onda, usualmente denominado blanco de referencia. Por lo tanto, en la práctica, la medición de la reflectancia de la muestra es relativa en comparación con un reflector no estándar. La relación es:

$$\text{Reflectancia} = \frac{\text{Intensidad de la reflectancia de la muestra}}{\text{Intensidad de la reflectancia de referencia}}$$

La identificación de minerales por espectroscopía NIR se basa en la comparación de un espectro de muestra con espectros de referencia de análisis de minerales individuales.

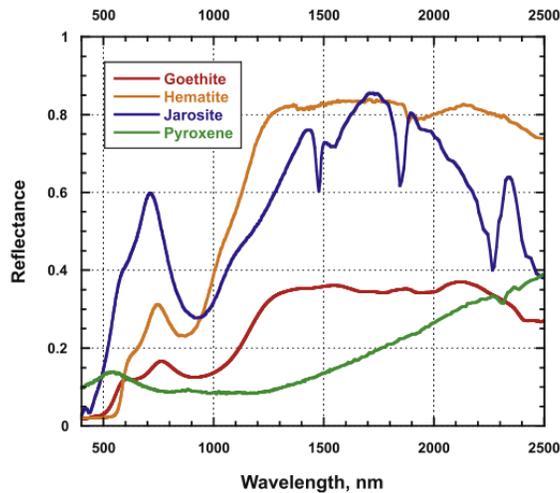


Figura 3 Ejemplo de Reflectancia espectral para cuatro minerales de hierro. tomado de (Bioucas-Dias José et al, 2012)

Specmin®Pro y TSG (the Spectral Geologist) son softwares comerciales que poseen librerías de espectros comúnmente utilizados para la identificación de minerales.

3.2.2 Quantitative Evaluation of Materials by Scanning (QEMSCAN)

QEMSCAN es un sistema de análisis de minerales basado en microscopio electrónico de barrido que utiliza un detector de electrones retrodispersados (EEB) para delinear partículas de resina epoxi, utilizando el nivel de brillo, antes de que se recopilen los espectros de rayos X, generalmente en una cuadrícula, con datos mapeados en cada píxel o punto.

Los espectros se comparan con una biblioteca de referencia de minerales (conocida como Protocolo de Identificación de Especies o SIP) y cada punto se identifica automáticamente en función de la micro química. El brillo de la EEB también puede incluirse en los criterios de identificación. La información de densidad mineral se aplica luego para cuantificar la abundancia mineral general en porcentaje en peso. Tanto el contenido mineral como la textura pueden investigarse mediante esta técnica.

QEMSCAN produce una mineralogía precisa para las fases objetivo, así como la gran mayoría de las gangas sulfuradas y no sulfuradas desde la traza hasta la mayor abundancia, aunque las bajas abundantes requerirán el análisis de un mayor número de partículas (y, por lo tanto, bloques) para mantener niveles de error aceptables.

El isomorfismo se denomina en mineralogía cuando dos minerales son iguales en composición química, pero su estructura molecular es diferente. Para ilustrar esto, el aragonito Co_3Ca tiene una estructura rómbica, mientras que la calcita Co_3Ca tiene una estructura trigonal. QEMSCAN tiene limitaciones para la identificación de minerales con isomorfismo.

Los minerales con una química similar pero en diferentes estados de hidratación también pueden ser difíciles de separar. La moscovita y la illita son minerales comunes en los sistemas de pórfido de cobre que demuestran este problema.

Finalmente, muchos de estos problemas pueden abordarse, y a menudo resolverse, en el paquete de software iDiscover a través de técnicas de procesamiento de datos, como procesadores de fase límite, agrupaciones minerales, etc.

3.2.3 X-Ray Diffraction (XRD)

La difracción de rayos X o DRX se usa ampliamente para la identificación de minerales cristalinos. La muestra se procesa en forma de polvo que consiste en granos finamente molidos de material a estudiar. Los rayos X se generan al bombardear un objetivo metálico, a menudo Cu, con un haz de electrones emitidos por un filamento de tungsteno. Durante el análisis, se registra un patrón de difracción 2D, ya que el detector del sistema gira a través de un ángulo (2θ), produciendo una

serie de picos correspondientes a los distintos espacios "d" (o espacios atómicos) en la red cristalina de un mineral.

Las posiciones y las intensidades de los picos se utilizan para identificar la estructura / mineralogía subyacente de la muestra.

La mayor fortaleza de XRD es la capacidad de identificar la gran mayoría de las especies de ganga, en particular los minerales arcillosos. Los minerales amorfos o poco cristalinos pueden ser imposibles o difíciles de detectar; la agudeza máxima puede ser una indicación del grado de cristalinidad, y la agudeza disminuye a medida que un mineral se vuelve más amorfo. La calidad de los resultados finales también puede verse limitada por la habilidad y la experiencia del operador que interpreta los patrones de difracción.

Muchos picos se superponen, lo que puede causar desafíos al identificar adecuadamente las diferentes fases minerales. Por ejemplo, la clorita y la esmectita son un desafío común. En este caso, se puede utilizar trabajo adicional que implica la separación de arcilla, seguido de la adición de glicol y tratamiento térmico. La orientación preferencial de los filosilicatos también puede sesgar los resultados.

El software, como TOPAS o SIROQUANT, puede ejecutar en el patrón de difracción para proporcionar estimaciones de abundancia mineral de las fases principales que se identifican.

3.3 Geoestadística

La geoestadística es una rama de la estadística que trata fenómenos espaciales, siendo el interés principal la estimación y la simulación de dichos fenómenos. Los orígenes de la geoestadística están asociados a la minería, donde se suele citar los trabajos de Sichel en 1947 y Krige en 1951. La idea básica es representar cualquier valor desconocido "Z" como una variable aleatoria Z, para poder encontrar una distribución de probabilidad que permita modelar la incerteza de z.

3.3.1 Definición de conceptos fundamentales

Variable aleatoria (Z): Es por lo general espacialmente dependiente por lo que se usa la notación $Z(u)$, donde u corresponde al vector de coordenadas. Z es también dependiente de la información, entendiéndose por esto a que su función de distribución cambia a medida que más información del valor desconocido $z(u)$ se encuentra disponible.

Función aleatoria (Z(u)): Una función aleatoria es un conjunto de variables aleatorias definidas sobre un campo de interés, o sea, $\{Z(u); u \in \text{area de estudio}\}$. Una función aleatoria es denotada, al igual que una variable aleatoria, como $Z(u)$. Algunos sinónimos utilizados son campo aleatorio, proceso aleatorio o estocástico.

Distribución espacial: una función aleatoria es representada por un conjunto de K funciones de distribución acumulada (cdf), para cualquier número K y para cualquier elección de K posiciones $u_k; k = 1, \dots, K$:

$$F(u_1, \dots, u_k; z_1, \dots, z_k) = \text{Prob} \{Z(u_1) \leq z_1, \dots, Z(u_k) \leq z_k\}$$

La función de distribución acumulada (cdf) de una variable aleatoria continua es denotada como:

$$F(u; z) = \text{Prob} \{Z(u) \leq z\}$$

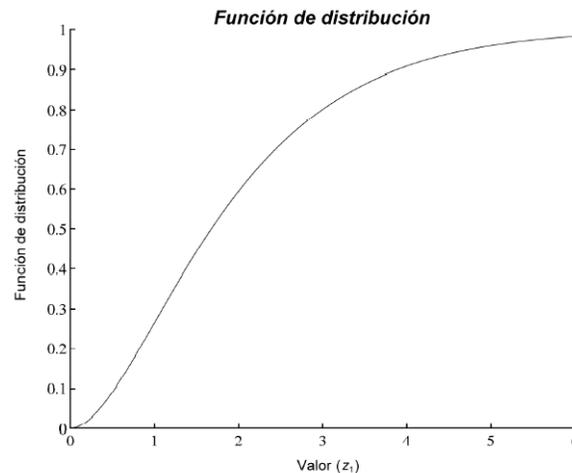


Figura 4 Ejemplo de función de distribución acumulada (cdf). Fuente Apunte de Geoestadística -Emery

Cuando la cdf es condicionada a un conjunto de datos consistentes de n valores vecinos, $(Z(u_\alpha) = z(u_\alpha), \alpha = 1, \dots, n)$, se usa la notación “condicional a n”, lo que define una función de distribución acumulativa condicional (ccdf):

$$F(u; z | n) = \text{Prob} \{Z(u) \leq z | n\}$$

Es importante notar que la ccdf $F(u; z | n)$ depende de la posición u, del tamaño del conjunto de datos condicionantes, de la configuración geométrica de estos (posición de los *datos* $u_\alpha, \alpha = 1, \dots, n$) y de sus valores $z(u_\alpha)$.

La **estacionaridad** supone que los valores que se encuentran en las diferentes regiones del campo presentan las mismas características y, por ende, pueden considerarse como distintas realizaciones del mismo proceso aleatorio. Es decir, las distribuciones de la función aleatoria $\{Z(u); u \in A\}$ estacionaria dentro de un campo A son invariantes por traslación.

La hipótesis de estacionaridad facilita la inferencia. Por ejemplo la cdf:

$$F(z) = F(u, z), \forall u \in A$$

puede ser inferida a través de un histograma acumulado de los valores de la variable presentes en las distintas posiciones del campo A.

3.3.2 Variograma

El variograma es una medida de la variabilidad de una variable regionalizada a lo largo de una dirección del espacio. Esta consiste en la diferencia entre los valores de dos variables según aumenta la distancia entre los puntos. Bajo el supuesto de estacionaridad, el variograma se define como:

$$\gamma(h) = \frac{1}{2} \cdot E\{ (Z(u+h) - Z(u))^2 \}$$

Se estima partir de los pares de datos cuya separación es igual (o casi igual) a un vector h dado. El estimador del variograma se conoce como 'variograma experimental'. Los parámetros que se deben definir para el cálculo del variograma experimental son:

- Dirección de interés: acimut, inclinación
- Distancias de interés: paso, número de pasos
- Tolerancia en la dirección: tolerancias angulares, anchos de banda
- Tolerancia en las distancias.

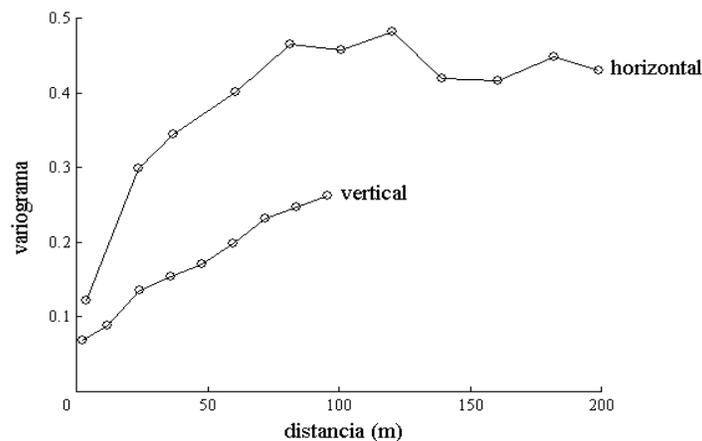


Figura 5 Ejemplo de variograma experimental en direcciones horizontal y vertical. FUENTE: (MI68A -Emery)

El variograma experimental es un estimador insesgado del variograma teórico, pero es poco robusto, en especial cuando

- el número de datos es pequeño
- la distancia |h| considerada es grande

- la malla de muestreo es muy irregular o preferencial
- la distribución de los valores es muy asimétrica o presenta valores extremos.

El variograma experimental requiere ser modelado para poder ser utilizado posteriormente en los procesos de estimación (kriging o cokriging) y simulación.

En el caso multivariable, se define un conjunto de variogramas directos (propios a cada variable) y cruzados, los cuales miden la interdependencia entre variables.

3.3.3 Kriging

El kriging es una colección generalizada de técnicas de regresión lineal para minimizar una varianza de estimación definida a partir de un modelo previo de variograma.

La resolución de un problema de estimación por kriging consiste en una combinación lineal ponderada (promedio ponderado) de los datos:

$$Z^*(u) = a + \sum_{i=1}^n \lambda_i Z(u_i)$$

donde $Z^*(u)$ es el valor estimado para la posición u , $\{Z(u_i), i=1\dots n\}$ son los valores de los datos en las posiciones $\{u_i, i=1\dots n\}$, a es un coeficiente aditivo y $\{\lambda_i, i=1\dots n\}$ son ponderadores. Estos ponderadores dependen de distancia entre los datos y la continuidad espacial de la función aleatoria $Z(u)$, reflejada por su variograma.

Por ejemplo, si el variograma es regular se privilegian los datos cercanos, mientras que en casos de existir efecto pepita la ponderación se reparte entre los datos.

Las limitaciones del kriging dice relación al suavizamiento, es decir, los valores estimados son menos dispersos que los valores verdaderos, por lo que no se puede predecir la ocurrencia de valores extremos.

La técnica de kriging se generaliza a la estimación de varias variables usando sus variogramas directos y cruzados, con la llamada técnica de cokriging (Wackernagel, 2003).

3.3.4 Kriging Ordinario

El kriging ordinario asume que la media poblacional de la variable regionalizada es desconocida, es decir, m es desconocida para esta variable regionalizada $E[Z^*(u)]$.

$$Z^*(u) = \sum_{i=1}^n \lambda_i Z(u_i)$$

La función objetivo de este kriging ordinario debe minimizar la varianza del error cuadrático entre el valor real y el valor estimado, de decir, se debe minimizar la siguiente expresión. $E[Z^*(u) - Z(u)]^2$. Por otro lado, se debe cumplir que $\sum_{i=1}^n \lambda_i = 1$, restricción que garantiza que la predicción sea insesgada, es decir, la diferencia entre la media aritmética del valor real y estimado sea cero.

El vector de ponderadores que cumple las restricciones antes mencionadas para una variable regionalizada, es el resultado de la resolución del sistema de ecuaciones y que satisface las restricciones.

$$\begin{pmatrix} C(x_1 - x_1) & \dots & C(x_1 - x_n) & 1 \\ \vdots & & \vdots & \vdots \\ C(x_n - x_1) & \dots & C(x_n - x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \\ 1 \end{pmatrix}$$

Donde:

$C(x_j - x_i)$ = es la covarianza entre $Z(u_j)$ y $Z(u_i)$ de la variable regionalizada.

μ = parámetro de Lagrange que permite cumplir la restricción de $\sum_{i=1}^n \lambda_i = 1$

λ_i = ponderadores para la asignación del estimador $Z^*(u)$ y/o variable regionalizada

$$C_{ij} = cov(z(u_j), z(u_i)) = C(0) - \gamma(u_i - u_j)$$

En ambos casos, la matriz del lado izquierdo representa la relación entre los datos conocidos mediante las correlaciones que sólo dependen de la distancia relativas, mientras que el vector del lado derecho representa la relación entre cada dato y el punto a estimar mediante la correlación que depende de la posición relativa entre el dato conocido y la ubicación del punto o bloque a estimar, según corresponda. Además, por la hipótesis de estacionaridad de segundo orden, es posible trabajar con los variogramas en forma equivalente a las covarianzas, haciéndose útil el modelamiento vario gráfico.

3.3.5 Cokriging Ordinario

El cokriging ordinario es la extensión del kriging ordinario al caso multivariable. Su objetivo es estimar la variable en un punto del espacio a partir de las mediciones de ésta y de una o varias variables adicionales correlacionadas ubicadas en sitios circundantes. La estructura espacial de las muestras y su configuración geométrica son consideradas mediante el uso del variograma cruzado.

Se introduce una variable auxiliar Y, como una medición adicional a la variable regionalizada Z. esta puede ser re-escrita como:

$$Z^*(u) = \sum_{i=1}^{nz} \lambda_i Z(u_i) + \sum_{i=1}^{ny} \alpha_i Y(u_i)$$

Donde:

$z^*(u)$ = variable regionalizada que deseamos estimar

$Z(u_i)$ = vector espacial multivariado medido en n_z registros $i=1, \dots, n_z$

$Y(u_i)$ = vector espacial multivariado adicional medido en n_y registros $i=1, \dots, n_y$.

Las siguientes dos restricciones a los ponderadores son impuestas para estar libres de sesgo, $\sum_{i=1}^{n_z} \lambda_i = 1$ con $\sum_{i=1}^{n_y} \alpha_i = 0$. Finalmente, el estimador puede ser escrito como $Z^*(u) = \{z_1^*(u), z_2^*(u), \dots, z_k^*(u), \dots, z_{n_z}^*(u)\}$, como también de forma matricial.

$$\begin{pmatrix} C(z_1 - z_1) & C(z_1 - y_1) & 1 & 0 \\ C(y_1 - z_1) & C(y_1 - y_1) & 0 & 1 \\ 1' & 0' & 0 & 0 \\ 0 & 1' & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \alpha \\ \mu_z \\ \mu_y \end{pmatrix} = \begin{pmatrix} C(z_1 - x_0) \\ C(y_1 - x_0) \\ 1 \\ 0 \end{pmatrix}$$

Las ventajas del cokriging son la utilización de la variable auxiliar y esta correlacionada con la variable de interés, especialmente cuando se tiene altos costo de muestreo.

3.4 Concentración de minerales

Los minerales son sustancias naturales inorgánicas con una posición química y estructura atómica definida, aunque el termino mineral es comúnmente usado en un sentido más extenso que incluye algún valor económico que se extrae de la tierra. El valor económico de los metales es gobernado principalmente por la oferta y la demanda. La oferta incluye los volúmenes de extracción actuales, nuevos depósitos descubiertos y el reciclaje.

El procesamiento de minerales comprende dos etapas: reducción para liberar los granos de mineral de interés, y la separación física de las partículas de interés del resto. La liberación de minerales de interés a través de la reducción de tamaño o conminución, que envuelven las etapas de chancado y de molienda. Una vez liberadas las partículas de interés, estas deben ser separadas de las partículas de ganga y estas se realizan aprovechando las propiedades de los diferentes minerales.

Las propiedades físicas que son usadas para concentrar minerales son: ópticas, densidad, propiedades de superficie, susceptibilidad magnética, conductividad eléctrica.

La concentración por flotación de espuma se apoya en las propiedades de superficie para concentrar los minerales de interés.

3.4.1 Concentración por Flotación de minerales

La flotación se define como un proceso químico de tensión superficial que separa ganga de compuestos de interés para permitir una extracción económica. El mineral molido se adhiere

superficialmente a burbujas de aire previamente insufladas, lo que determina la separación de los minerales de interés.

Los minerales son clasificados en tipos polar o no polares de acuerdo a las características de su superficie. Las superficies de minerales no polares se caracterizan por enlaces moleculares relativamente débiles, es decir, las superficies no polares no se adhieren a los dipolos de agua, y en consecuencia son naturalmente hidrofóbicos. Minerales de este tipo, son grafito, sulfuros, molibdenita, diamantes, carbón y talco (Wills, 2006).

Minerales con fuerte enlace covalente o superficie iónica son conocidos como tipo polar y reaccionan fuertemente con moléculas de agua, estos minerales son naturalmente hidrofílicos.

3.4.2 Colectores

La adhesión del mineral a una burbuja de aire dependerá de las propiedades hidrofílicas y aerofílicas de cada especie mineral.

Los colectores son compuestos orgánicos que aumentan la capacidad aerofílica de los minerales a través de la absorción de su superficie y reduciendo la estabilidad de una capa hidratada, es decir, alterando su capacidad hidrofóbica. Los colectores pueden ser compuestos non-ionizada, que son prácticamente insolubles y fuertemente hidrofóbicos.

Los colectores son agregados al material molido con agua denominado pulpa y su aplicación se realiza por un tiempo que permita adsorción durante la etapa de agitación, conocido como periodo de condicionamiento.

Un colector ionizado puede ser clasificado según tipo de ion: aniónico o catiónico. Los colectores aniónicos se componen de dos grupos de acuerdo a la estructura polar: tipo “sulfhydryl” y tipo “oxyhydryl”. La Figura 6 muestra una figura esquemática con los tipos colectores.

3.4.3 Espumantes

Los espumantes tienen por objetivo durante la concentración por flotación ayudar a la formación de burbujas pequeñas, lo que aumenta las colisiones de partículas con burbujas porque aumenta el área superficial disponible a través de burbujas pequeñas y por lo tanto aumenta la cinética de flotación.

Los espumantes buscan reducir la velocidad de subida de la burbuja, lo que aumenta los tiempos de residencia de las burbujas en la pulpa, y con esto el número de colisiones.

Los espumantes fomentan la formación de espuma que reduce el estallido de burbujas cuando estas alcanzan la parte superior de la pulpa, permitiendo recolectar estas sobre flujo como un producto de la flotación.

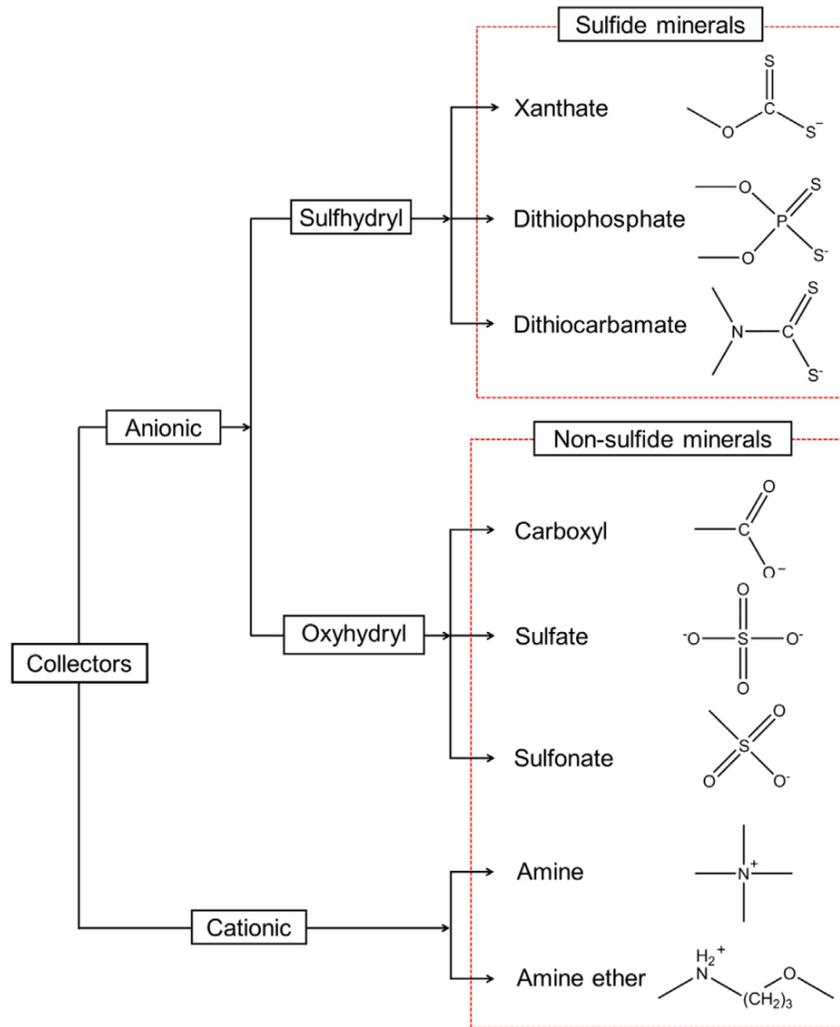


Figura 6 Clasificación de colectores ionizantes de acuerdo al tipo anión, catión y sus aplicación (tomado de (Wills, 2006))

3.4.4 Reguladores

Reguladores o modificadores son usados extensivamente en flotación para modificar la acción de los colectores, sea está intensificando o reduciendo el efecto repelente del agua en la superficie de mineral, como también volviendo la acción del colector más selectiva.

Los reguladores pueden ser clasificados como activadores, depresores, dispersantes o modificadores de pH.

3.4.5 Ensayos de flotación

Las pruebas de laboratorio son desarrolladas para determinar el desempeño de un mineral a las opciones de reactivos disponibles para un mineral específico y el tamaño de la planta para un determinado rendimiento.

Los ensayos de flotación también son llevados a cabo en minerales con muestras representativas, o muestras de plantas existentes extraídas durante el circuito, con el objetivo de mejorar los procedimientos y/o experimentar nuevos reactivos.

Las variaciones mineralógicas existente en un yacimiento, ha requerido elaborar procedimientos de identificación de minerales, muestra representativa y caracterización de dominio de geo metalurgia, que permitan explicar estas variaciones y anticipárseles.

3.4.5.1 Ensayos de flotación discreta o Batch

La mayor parte de las pruebas de flotación se realizan en celdas utilizando lotes de muestras usualmente con masas entre 500 a 1000 gr. Las celdas mecánicamente agitadas donde la velocidad de rotación de los impulsores puede ser ajustado para cada prueba de manera independiente.

La corriente de aire es impulsada desde la parte inferior generando burbujas finas que atrapan las partículas de mineral recogidas y eliminadas como una espuma mineralizada en la parte superior.

Las pruebas por lotes son bastante sencillas, aunque se pueden complejizar para emular procesos en múltiples etapas. El ensayo de ciclo cerrado se denomina a múltiples etapas de flotación batch para medir el efecto de cargas circulantes y proveer un mejor desempeño de procesamiento de minerales en varias etapas.

La máquina de flotación de laboratorio Denver es popularmente utilizada en la industria minera. Se compone de un brazo mecánico móvil de flotación suspendido para subir y bajar los impulsores de burbujas y difusores de burbujas. Existe una variedad de tanques de diferentes tamaños desde 250 gr a 2 Kg.



Figura 7 Máquina de flotación de laboratorio Denver Metso D-12

4 METODOLOGÍA PROPUESTA

La implementación de una metodología de estimación 3D con el porcentaje de recuperación de cobre (%RecCu) propuesto en este estudio, establece dos etapas para la utilización de datos espectrales:

1.- El objetivo de la primera etapa es evaluar dos metodologías para predecir la recuperación de cobre a través de técnicas de regresión. El entrenamiento supervisado con datos de espectrometría incorpora desafíos asociados a la calidad de los datos, sean estos el ruido y/o información redundante. La eliminación de espectros anómalos se incorpora como una etapa temprana.

La calibración y validación del modelo propuesto fue realizada y la valoración del desempeño se revisó con información complementaria QEMSCAN.

2.- El objetivo de la segunda etapa es incorporar la información de flotación y la espectral en un modelo tridimensional. El desafío de integrar estos datos se debe a la asimetría en volumen de información, sea esta en número de datos disponibles como en su complejidad.

Este análisis propone la construcción de una base de datos integrada con una reducción de la información espectral a través de componentes principales. Una vez interpolados las componentes principales y los datos de recuperación, permitirá evaluar dos alternativas de procesar la información, la primera propone usar directamente de la interpolación (método directo), mientras que la segunda propone un procesamiento de los espectros interpolados usando técnicas de regresión (método indirecto).

4.1 Etapa 1.- Regresión Multivariable

- a) **Construir una base de datos integrada:** Los resultados de ensayos de flotación *rougher* y lecturas de espectrómetro fueron unidos como una base de datos heterotópica. Esta base de datos se construye con una muestra única que obtiene tres resultados: flotación, espectrometría y QEMSCAN.
- b) **Descartar datos anómalos:** Un análisis exploratorio de ensayos de flotación *rougher* y de lecturas de espectrómetro anómalos fue realizado para descartar inconsistencias.
- c) **Implementar técnicas de regresión:** PCA y PLSR fueron utilizados para predecir el desempeño de flotación *rougher* para %RecCu.
 - a. Calibrar el modelo
 - b. Testear o evaluar el desempeño del modelo.
- d) **Evaluar el desempeño de las técnicas de regresión:** Similar al concepto de validación cruzada se evalúa el resultado de los errores.

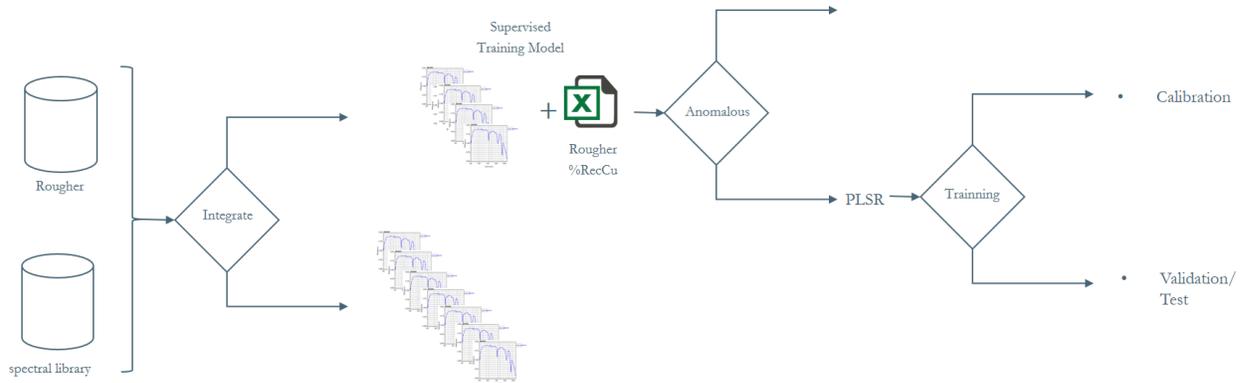


Figura 8 Esquema conceptual de integración de información de espectrómetros y datos de flotación de cobre.

4.2 Etapa 2.- Predicción de un modelo 3D recuperación de cobre

- a) **Incluir georreferencia a base datos de etapa 1 letra a):** la base de datos incorpora coordenadas para el proceso de interpolación del modelo 3D. Definir especificación de tamaño del modelo de bloques.
- b) **Reducir información espectral a través de selección de componentes principales:** esta etapa revisa el porcentaje de varianza explicado por las componentes.
- c) **Calcular y modelar el variograma experimental:** cálculo de variogramas experimentales entre % RecCu (flotación *rougher*) y componentes principales de información espectral.
- d) **Interpolar y reconstruir información espectral:** Ello a través de selección de componentes principales.
 - Propuesta 1: Estimaciones de %RecCu por cokriging, usando las componentes principales de datos de espectrómetros como covariables.
 - Propuesta 2: Estimaciones de %RecCu usando PLSR aplicado a las componentes principales de datos de espectrómetros estimadas por cokriging de la recuperación de cobre, pero reconstruidas como un dato espectral estimado en el bloque.

La Figura 9 presenta un esquema conceptual de la metodología de predicción de un modelo 3D de recuperación de cobre.

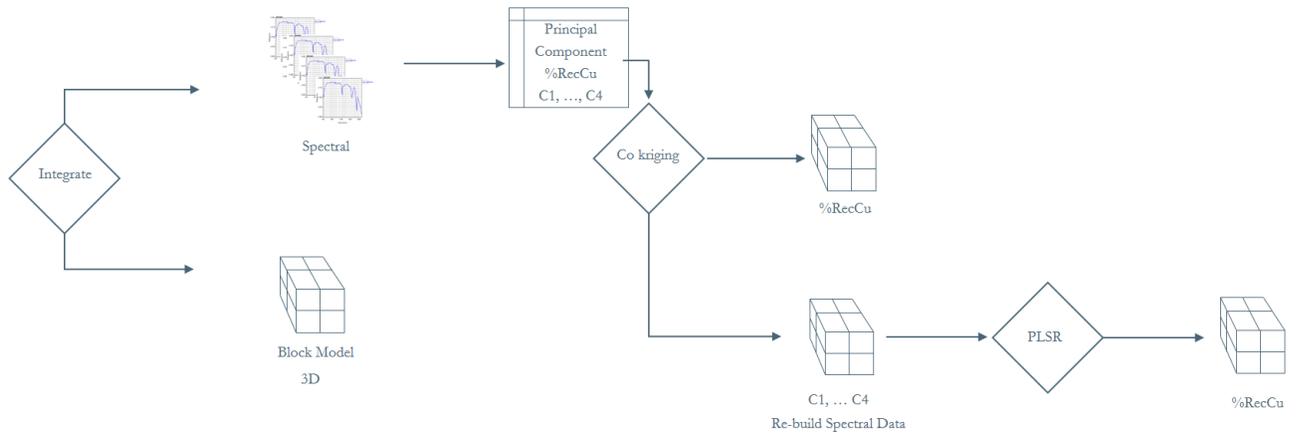


Figura 9 Esquema conceptual de integración de espectros al modelo 3D propuesta para predecir la recuperación de Cobre

5 CASO DE ESTUDIO

5.1 Etapa 1

5.1.1 Características generales de la base de datos integrada

Las pruebas de flotación *rougher* a escala de laboratorio fueron realizadas en celdas Denver con muestras 20% +100# Ty (P80 150 micrones). El procedimiento establecía una masa de alimentación de 1135 gr con 1135 cc de agua para luego ajustar el pH a 10. El concentrado de *rougher* se produce en seis etapas de 1, 2, 3, 6, 12 y 20 minutos. Un total de 650 muestras fueron utilizadas en este caso de estudio.

Tabla 1 Esquema de reactivos para ensayos de flotación

Etapa	Reactivos	Dosis (g/ton)
Molienda	D-101 Colector Primario	33
	SEC301 Colector Secundario	7
	Espumante - Frother P 749	25
Flotación	SEC-301 (agregado a los 7 minutos de iniciado el test)	7

La base de datos de espectrómetro está compuesta por los 650 registros que se suman a un total de 17,318 valores en el área de estudio. Las muestras fueron capturadas con el LabSpec en muestras de sondajes de 2 metros y 3 metros para aire reverso y diamantino, respectivamente. El tamaño de partícula es 10# Tyler y se exige un pasante del 90%.

Adicionalmente, La base de datos tiene resultados de QEMSCAN que fueron realizado a la cabeza de la masa utilizada para la etapa de flotación.

5.1.2 Revisión de datos anómalos

5.1.2.1 Datos anomalos de flotación de cobre

La ley de cobre promedio de alimentación fue 0.63%. La principal mena de cobre es calcopirita. Un 94% de los datos fue ensayado con 1135 gr, el grupo restante se realizó con 750 gr como parte de un cambio de procedimiento. La Tabla 2 y la Tabla 3 muestran un resumen de las estadísticas básicas de las leyes y masas de las pruebas de flotación *rougher*.

La Figura 10 muestra información de leyes de alimentación en (a), y la masa de los ensayos de flotación en (b). La Figura 11 muestra el resultados de la recuperación obtenidas de este ensayos.

Una revisión del balance de masa de la base de datos fue realizada, siendo la masa del concentrado el resultado de la suma de las seis etapas de flotación. No fueron identificados errores en la base de

datos. Los casos considerados anómalos fueron los registros #243, #34, #133, #9 y #266, con registro de masa de concentrado en doscientas veces sobre la media de 183 gr para ensayos de 1135 gr.

Una revisión de las recuperaciones para cada una de las seis etapas fue realizada. Ésta comprobó que el orden de magnitud de las masas fuera siempre creciente, es decir, concentrado de etapa uno es inferior al concentrado de la etapa 2 y así sucesivamente para las seis etapas.

No fueron identificados errores en la base de datos y adicionalmente, se eliminaron nueve ensayos que eran inconsistentes con la ley de alimentación y sus derivados.

Tabla 2 Estadísticas básicas de las leyes de cobre de los ensayos de flotación *rougher*

	Número	Media	Mínimo	Máximo	Std. Dev	C.V.
Ley Cu Concentrado	458	4.406	0.5877	19.468	2.5812	1.7
Ley Cu en Relave	458	0.073	0.006	0.61	0.0661	1.1
Ley Cu en Alimentación	458	0.634	0.139	3.24	0.3569	1.8

Tabla 3 Estadísticas básicas de masa de los ensayos de flotación *rougher*

	Número	Media	Mínimo	Máximo	Std. Dev	C.V.
Masa Concentrado	458	1107.6	729.6	1136.6	91.632	12.1
Masa Relave	458	947.5	514.7	1090	100.2516	9.5
Masa Alimentación	458	158.9	37.1	619.6	72.1639	2.2

La selección de muestras de los datos de calibración y testeo se realizó con un muestreo estratigráfico de los dos grupos de datos considerados con buen y mal desempeño metalúrgico. El umbral utilizado para definir buen y mal desempeño fue 91.3% RecCu, que corresponde a la media de los datos utilizados.

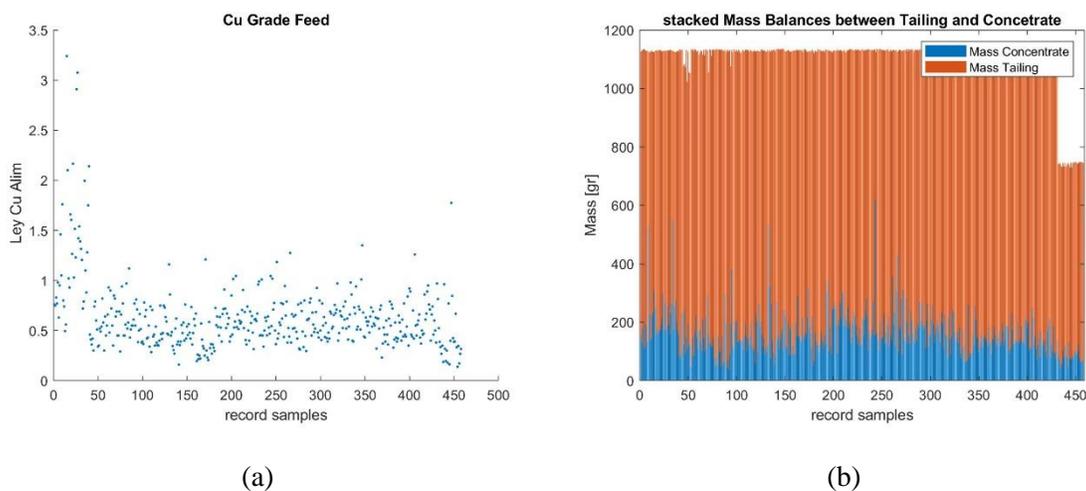


Figura 10 Muestra datos de entrenamiento en (a) %Cu en alimentación; (b) masa de alimentación.

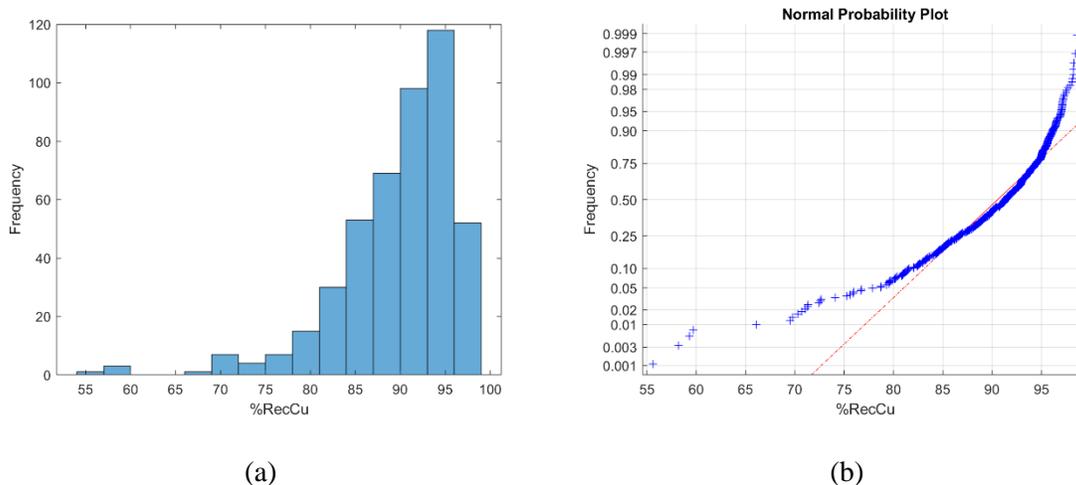


Figura 11 Muestra las características de %RecCu utilizado para la calibración del modelo: (a) histograma de %RecCu (b) gráfico de probabilidad normal de %RecCu.

5.1.2.2 Datos anómalos de espectrómetros

La identificación de datos espectrales anómalos se vuelve un trabajo altamente complejo. Las métricas propuestas en este análisis para la detección de espectros anómalos fueron la distancia de Cook y la distancia de Mahalanobis. Posteriormente, se contrastará con los errores residuales observados entre valores reales y estimados.

Una revisión de los espectros anómalos es presentada en la Figura 12. La distancia de Mahalanobis es la longitud entre un punto de datos y el centroide de un espacio multivariado (media global) (Mahalanobis, 1936). Para esto, el cálculo de la distancia toma en cuenta la covarianza entre las variables, situación que no ocurre con otras métricas.

Por ejemplo, esto no ocurre con una distancia euclidiana que solo mide una distancia geométrica de dos puntos, sin considerar la covarianza. En esta figura se aprecia que los casos #179, #271, #338, #180 y #34 son influyentes en las estimaciones de los coeficientes PLSR, mientras que los espectros influyentes para PCA son #447, #102, #101, #104 y #281. Por lo tanto, la intersección entre el grupo de espectros anómalos es nulo, es decir, la distancia propuesta para cada método fue distinta.

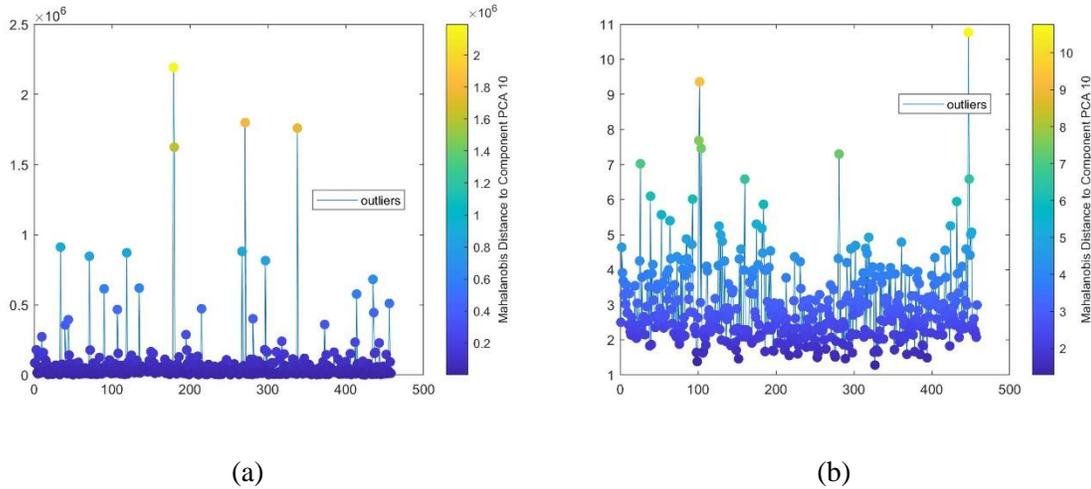


Figura 12 Distancias de Mahalanobis con 10 componentes. (a) PLSR obtenidas desde los datos de entrenamiento, (b) PCA.

La influencia de un espectro en la regresión lineal, sea esta PCA o PLSR, fue revisada con la distancia de Cook, esto propone evaluar de manera individual cada espectro y su impacto en la predicción, es decir, mide la influencia de un caso en todos los parámetros de estimados (Williams, 2016).

La distancia de Cook muestra la influencia de cada espectro (observación) en el ajuste de la %RecCu (variable respuesta). La

Figura 13 muestra que los espectros influyentes para distancia de Cook fueron #267, #28, #23, #6 y #otros 52 espectros.

La intersección entre ambas métricas muestra que #338 y #101 fueron revisados por su impacto en la predicción de %RecCu. La Figura 14 muestra los espectros mencionados como outliers según la distancia de Cook.

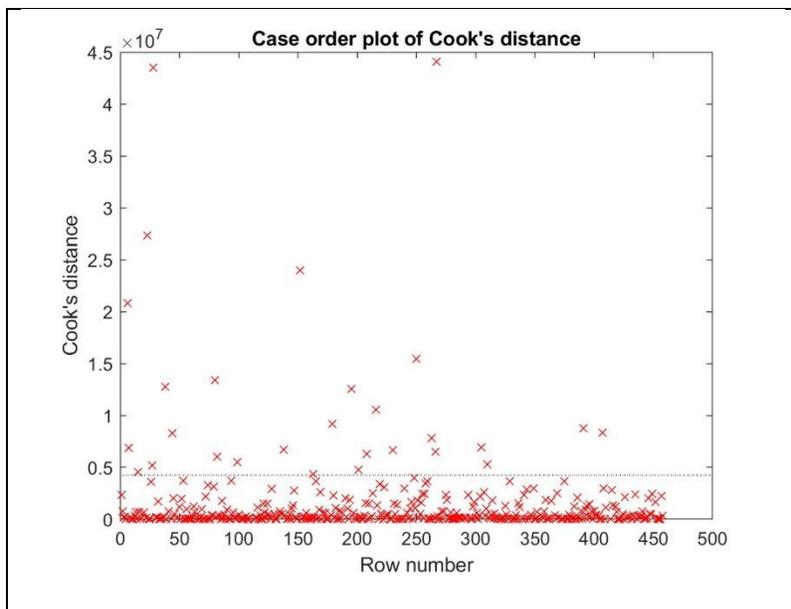


Figura 13 Distancia de Cook para el conjunto de entrenamiento

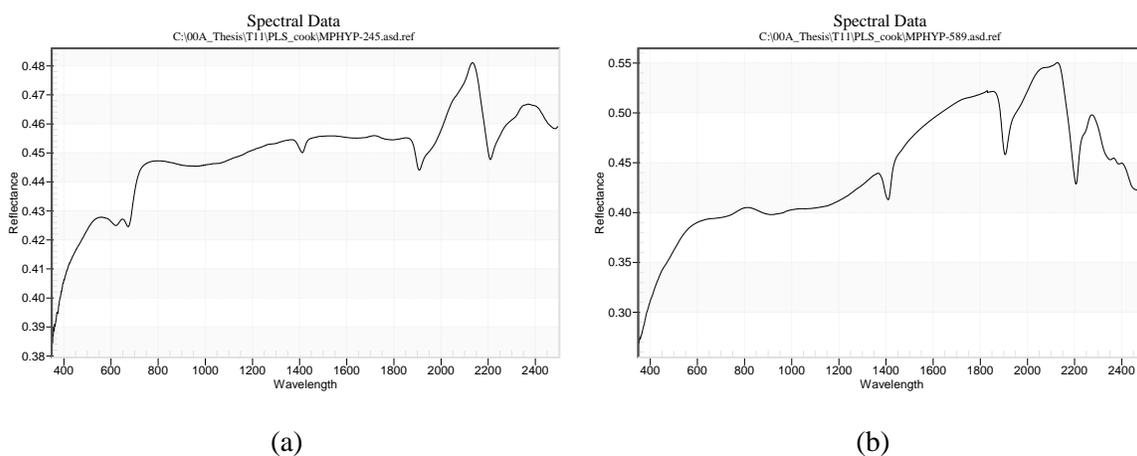


Figura 14 Muestra outliers sugeridos usando distancia de Cook (a) registro #308 (MPHY-589) con 58%RecCu y (b) registro #101 (MPHY-245) con 93%RecCu.

Una interpretación realizada en el software Specmin®Pro indica que el registro MPHY-245 es mayoritariamente montmorillonita, mientras que el registro MPHY-589 es una combinación entre caolinita (65%) e illita (35%). La Figura 15 muestra los resultados de QEMSCAN relativo a filosilicatos etiquetado como *QMS_Kaolinite_clays* que integra caolinita, illita y montmorillonita. La mena en ambos casos corresponde a calcosina, pero que en contenido de arcillas contiene el doble en proporción. Esto justificaría su mal desempeño en flotación con 58%RecCu.

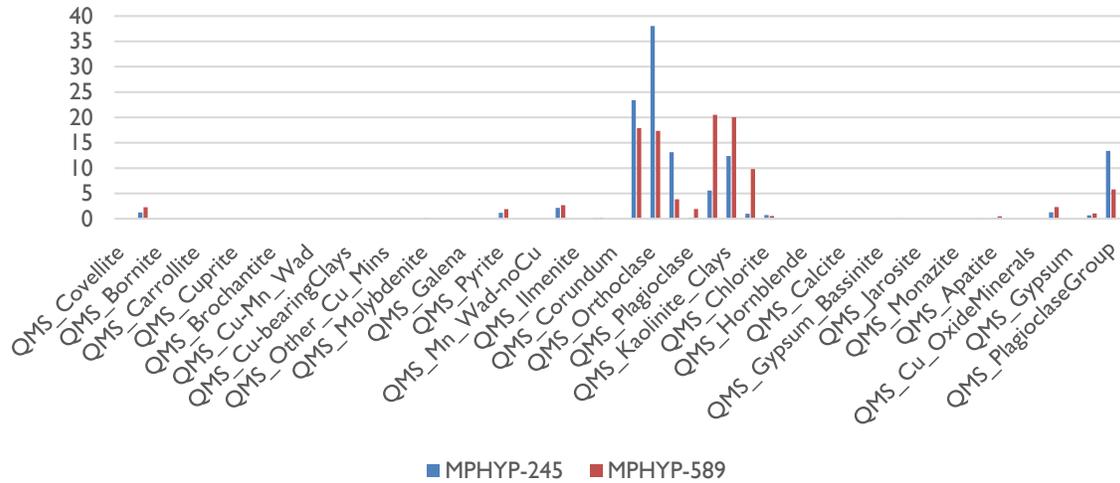
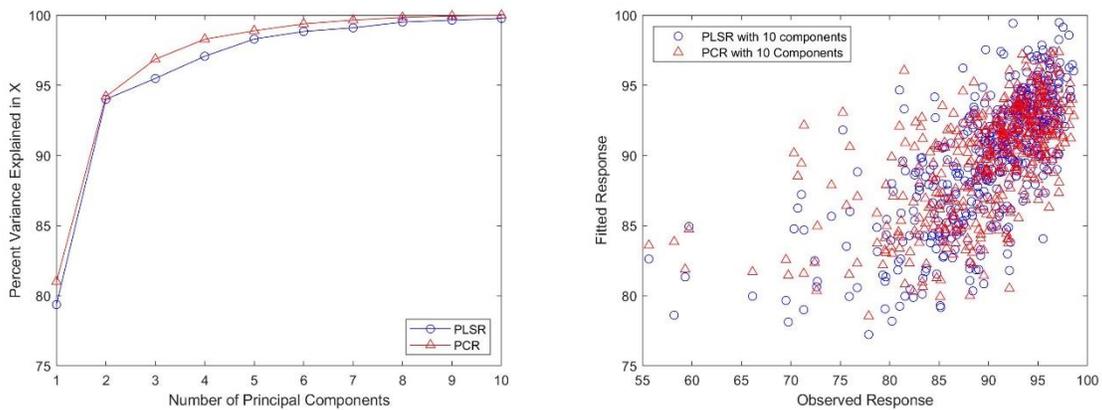


Figura 15 Cuadro comparativo de QEMSCAN para ambos registros con bajo desempeño MPHYP-245 y MPHYP-589.

5.1.3 Calibración

La etapa de calibración tiene por objetivo asegurar la calidad de los resultados. Una parte importante de esta etapa es la elección del número de componentes principales, lo que permitirá resumir la información espectral a un conjunto de datos de menor dimensión. El gráfico de la total varianza explicada fue utilizado para comparar ambos métodos de regresión usando componentes (PCR) y una regresión usando mínimos cuadrados parciales (PLSR). Durante la etapa de calibración se escogieron diez componentes principales para realizar las comparaciones.

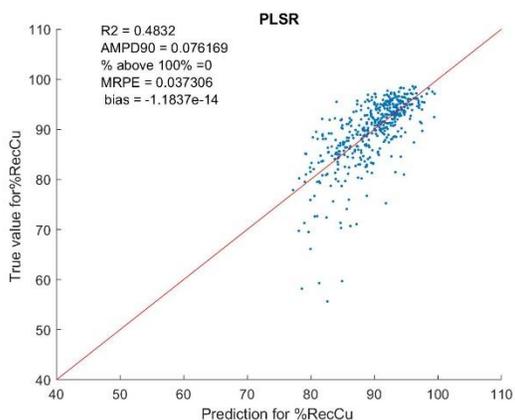
La capacidad de explicar la varianza de los datos para ambos métodos no muestra diferencias significativas. La Figura 16 indica la capacidad de ambas técnicas de reducir la variabilidad capturada en los espectros. PCR y PLSR no muestran diferencias significativas en la predicción de la variabilidad.



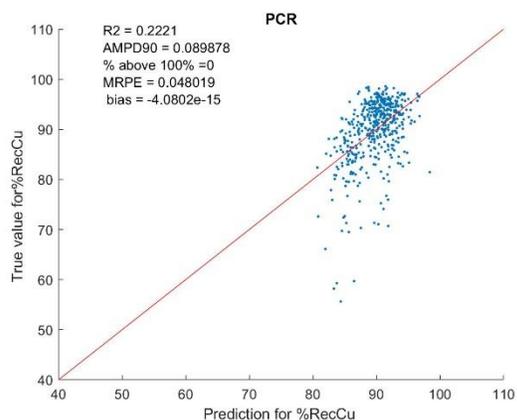
(a)

(b)

Figura 16 Comparación de métodos de regresión de datos de calibración de recuperación de cobre: (a) porcentaje de varianza explicado según número de componentes, (b) nube de correlación entre recuperación de cobre real y estimada.

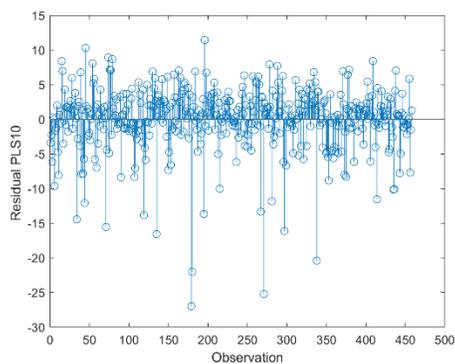


(a)

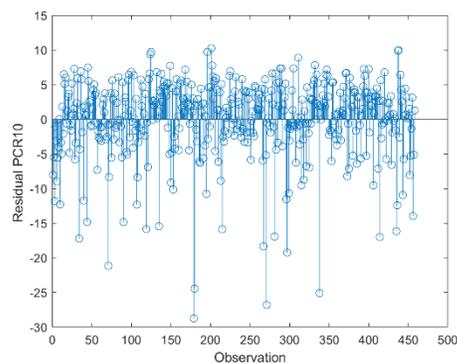


(b)

Figura 17 Scatter plot de los %RecCu reales versus los estimados usando (a) PLSR y (b) PCA



(a)



(b)

Figura 18 Errores residuales entre valores reales y valores estimados. En (a) para PLSR y en (b) para PCA

El coeficiente de determinación (R^2) muestra que una regresión de mínimos cuadrados parcial (PLSR) tendrá un mejor desempeño con un coeficiente de 0.4832 mientras que PCR tiene un R^2 de 0.2221, es decir, PLSR explica una mayor parte de la variabilidad de %RecCu a través del uso de espectros.

Este es un resultado esperable por que el algoritmo de PLSR considera la variable respuesta (%RecCu) durante la iteración de las variables latentes, y no así PCR que únicamente se enfoca en la varianza de las variables predictoras, que en este caso son los espectros.

Una segunda métrica utilizada en el ámbito de control de calidad (QAQC, Quality accuracy, Quality Control) es llamada AMPD90 (Absolute Differences Measure Pair Duplicated). Esta métrica indica el percentil 90 de la distribución de diferencias entre pares de datos real y estimado. PLSR obtiene 7.6%, menor al obtenido con PCR de 8.9%, es decir, PCR tiene diferencias entre real y estimado superiores a las obtenidas con el PLSR.

El RMSEP (Root Mean Squared Error of Prediction) muestra errores aleatorios menores en PLSR que PCR. La Figura 17 registra algunos resultados que permiten comparar entre ambos métodos.

En términos generales, PLSR presenta mejores indicadores de nivel de errores presentados, porque las métricas presentadas reflejan bajo nivel de sesgo, errores aleatorios y mejor explicación de la variabilidad de %RecCu.

Los errores residuales pueden ser divididos en sobre-estimación y sub-estimación. PLSR muestra que los espectros #196, #45, #74, #79, #409 presentan diferencias desde 8.3 puntos entre real y estimado (sub-estimación), del mismo modo PCR muestra que los espectros #201, #438, #437, #196, #125 presentan diferencia desde 9.7 puntos entre real y estimado (sub-estimación).

Los métodos tuvieron predicción inconsistente en un conjunto de espectros, pero sin intersección entre estos.

Por otro lado, los siguientes espectros #179, #271, #180, #338, #135 presentan diferencia desde 16 puntos como una sobre-estimación utilizando PLSR, mientras que PCR son los siguientes espectros #179, #271, #338, #180, #71 con diferencias desde 21 puntos (sobre-estimación).

La Figura 18 muestra una coincidencia para ambos métodos para #179 (MPHYP-356.asd); #271 (MPHYP-496.asd); #180 (MPHYP-357.asd). Cabe destacar que los resultados de flotación de estos tres ensayos fueron los entre 55-59 %RecCu.

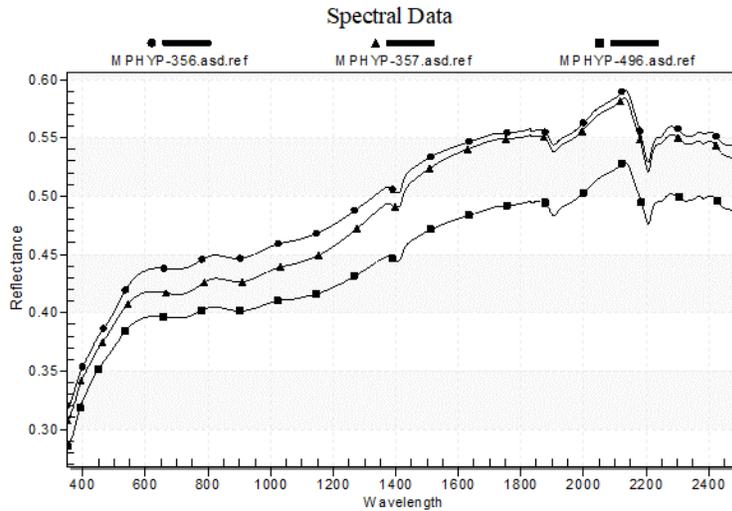


Figura 19 Muestra reflectancia NIR para los registros #179(MPHYP-356.asd); #271 (MPHYP-496.asd); #180(MPHYP-357.asd)

Una revisión de la información espectral usando Specmin®Pro muestra que los tres espectros tienen rasgos asociados a la presencia de montmorillonita, muscovita e illita. La Figura 19 muestra que estos tres espectros son muy similares, con leves diferencias en la magnitud de la reflectancia. Estas interpretaciones fueron confirmadas con los resultados QEMSCAN.

Los pesos asignados por PLSR en longitud de onda [1130,1470] y [2134,2195] corresponden a características del grupo hidroxilo, es decir, ion OH⁻, donde la mayoría de los filosilicatos contiene este ion OH⁻.

Recordemos las limitaciones en la interpretación de arcillas. Por ejemplo, la muscovita, la sericita y la illita suelen ser agrupadas en un mismo campo. Similar es el caso de caolinita, montmorillonita y esmecticas.

Los pesos asignados por PLSR en longitud de onda [1130,1470] y [2134,2195] corresponden a características del grupo hidroxilo, es decir, ion OH⁻, donde la mayoría de los filosilicatos contiene este ion OH⁻.

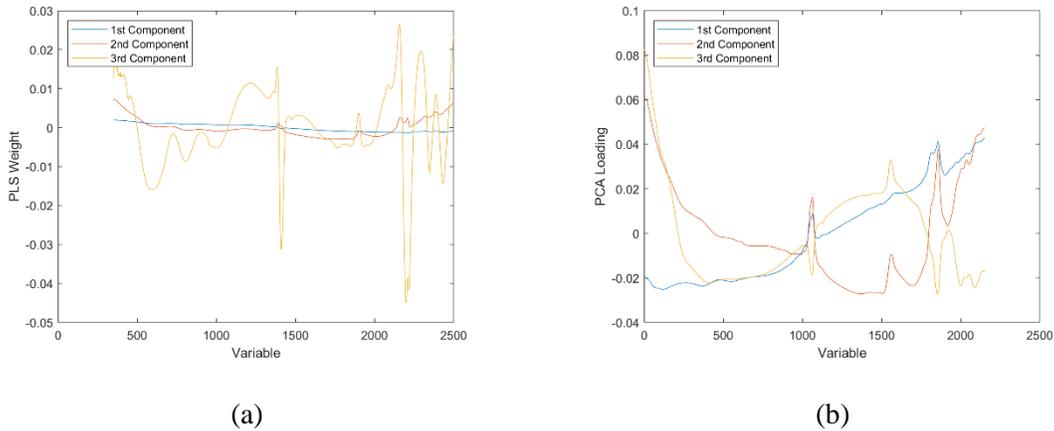


Figura 20 Muestra la asignación de los ponderadores para la generación de variables.

5.1.4 Validación del modelo regresión propuesto

La validación corresponde a una medida para evaluar la calidad del modelo propuesto. La Figura 21 y la Figura 22 presentan los scatter-plots de las predicciones del porcentaje de recuperación de cobre hechas con los dos modelos analizados versus el valor real de recuperación de cobre ensayado.

Regresión usando mínimo cuadrados parciales muestra un mejor desempeño explicando mayor variabilidad de %RecCu que utilizando regresión de componentes principales.

Una revisión de los errores residuales muestra coincidencia en los espectros con sobre-estimación mayor a 13 puntos. Los espectros #140, #128, #34, #184 y #165 tienen mal desempeño de predicción cuando se compara con el valor ensayado para métodos de regresión PLSR y PCR. Por otro lado, PCR muestra una sub-estimación de 7 puntos de recuperación #59, #69, #130, #126 y #83, mientras que para PLSR los siguientes espectros #59, #63, #7, #130 y #97.

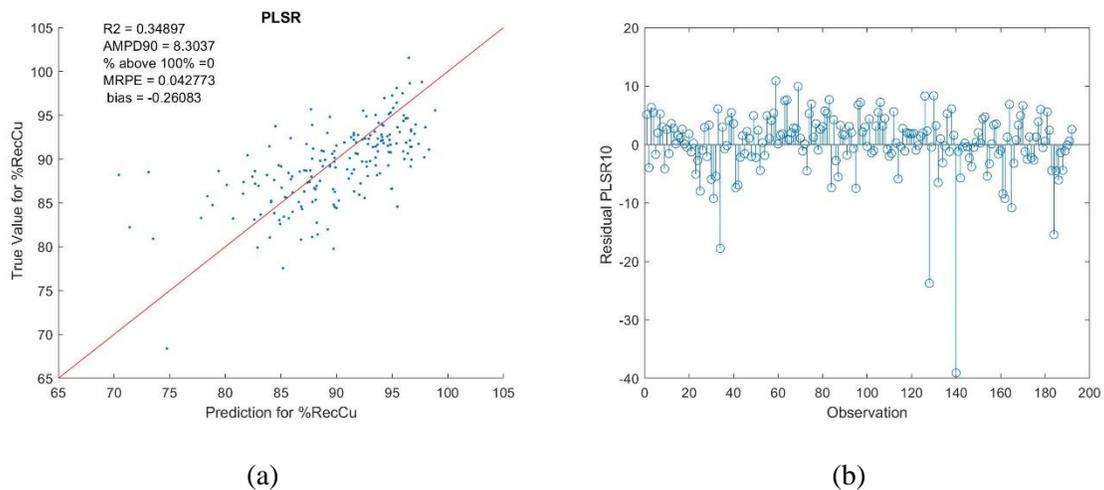
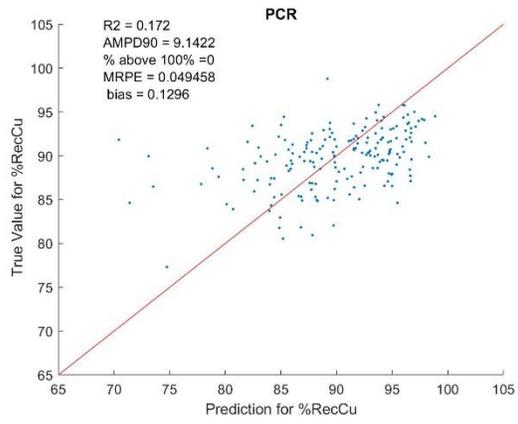
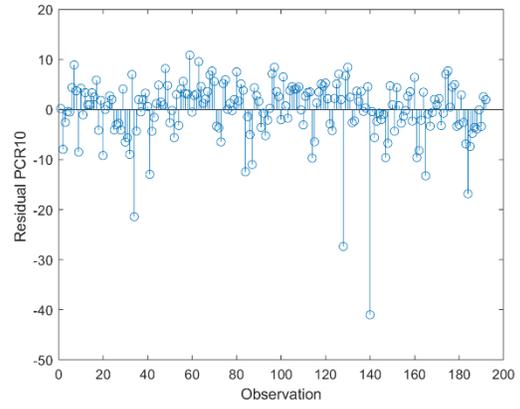


Figura 21 (a) Gráfico de dispersión del %RecCu usando PLSR. (b) Errores residuales las predicciones propuestas.



(a)



(b)

Figura 22 (a) Gráfico de dispersión del %RecCu usando PCR. (b) Errores residuales las predicciones propuestas.

5.2 Etapa 2 - Predicción de un modelo 3D de recuperación de cobre

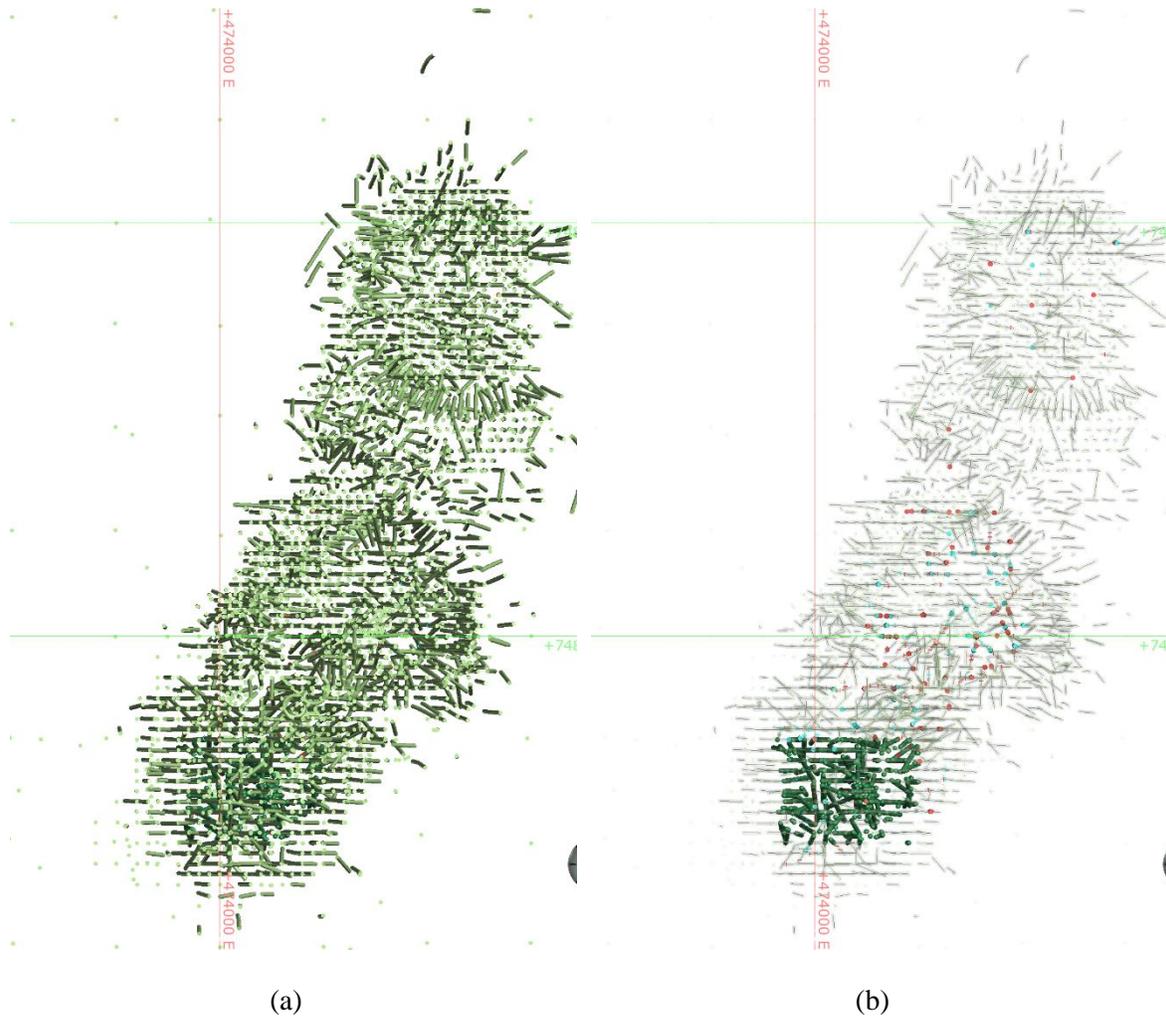


Figura 23 Vista en planta de la información del depósito. (a) Información espectral disponible. (b) Zona de trabajo para el modelo 3D.

Las muestras seleccionadas para el caso de estudio corresponden al sur de depósito, escogiéndose esta zona por tener buena densidad de datos, número suficiente de pruebas de *rougher* (670 ensayos), bien distribuidas entre buen desempeño y mal desempeño en términos de recuperación de cobre, es decir, cercana a una recuperación del 100% entre alimentado y recuperado.

La base de datos heterotópica se construyó con 17,318 espectros disponibles, el número disponible con ensayos de flotación *rougher* son 176 muestras y 61 de calibración.

La Figura 23 señala la ubicación del caso de estudio que se ubica en el sector sur del depósito, que tiene una mejor concentración de muestras.

El modelo de bloques está construido como un área regular de 2520 bloques, se utilizaron bloques regulares de 50 metros con la siguiente repetición de 14x10x18, en términos generales hablamos de un área de 700x500 metros.

5.2.1 Revisión de parámetros de cokriging

Las rutinas implementadas están escritas en el software Matlab.

El cálculo de variogramas experimentales fue omnidireccional, es decir una única dirección fue implementada. Para el cálculo de la variabilidad, la distancia de medición de variabilidad fue con pasos de 50 metros y se incluyó una tolerancia de 25 metros, con una distancia máxima de 500 metros, es decir, se utilizaron 10 pasos.

La naturaleza de los datos de reflectancia tiene registros muy cercanos a cero y la implementación de las rutinas incluye únicamente cuatro decimales, razón por la cual fueron amplificados en 1E6.

El modelamiento de los variogramas experimentales se realizó como dos estructuras anidadas, una esférica de alcance 30 m y otra exponencial de factor de escala 100 m, y un efecto pepita de 0.5 para todas las estructuras y se utilizó un ajuste de meseta automático. La utilización de ponderadores proporcional al número de datos fue descartado, como tampoco los datos fueron desagrupados.

La

Figura 25 muestra variogramas cruzados que fueron modelados en este estudio.

La estimación de cokriging considera un radio de búsqueda omnidireccional de 300 metros, no se dividió por octantes y se limitó el número óptimo de muestras en 200.

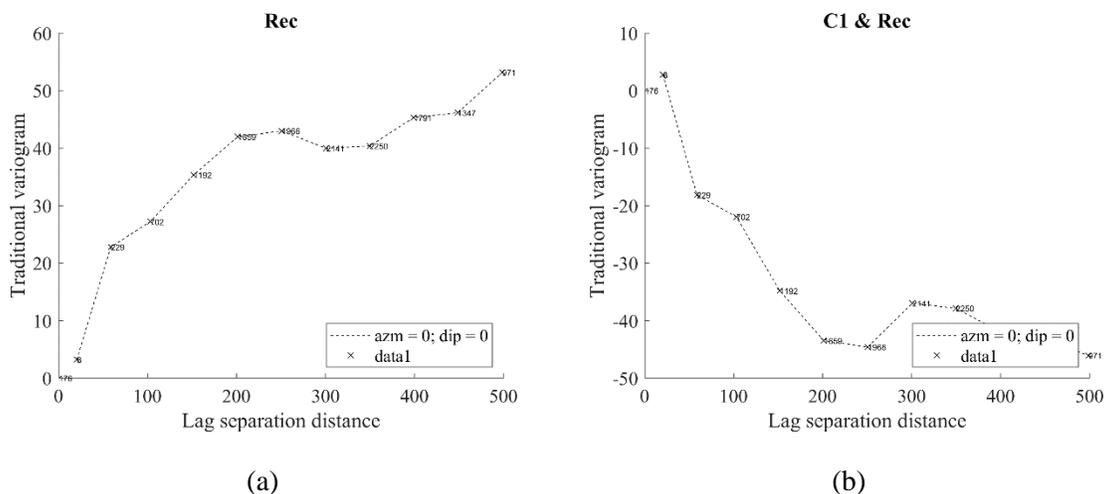


Figura 24 Muestra en (a) el variograma experimental directo de %RecCu y (b) el variograma experimental cruzado entre %RecCu y la componente principal 1.

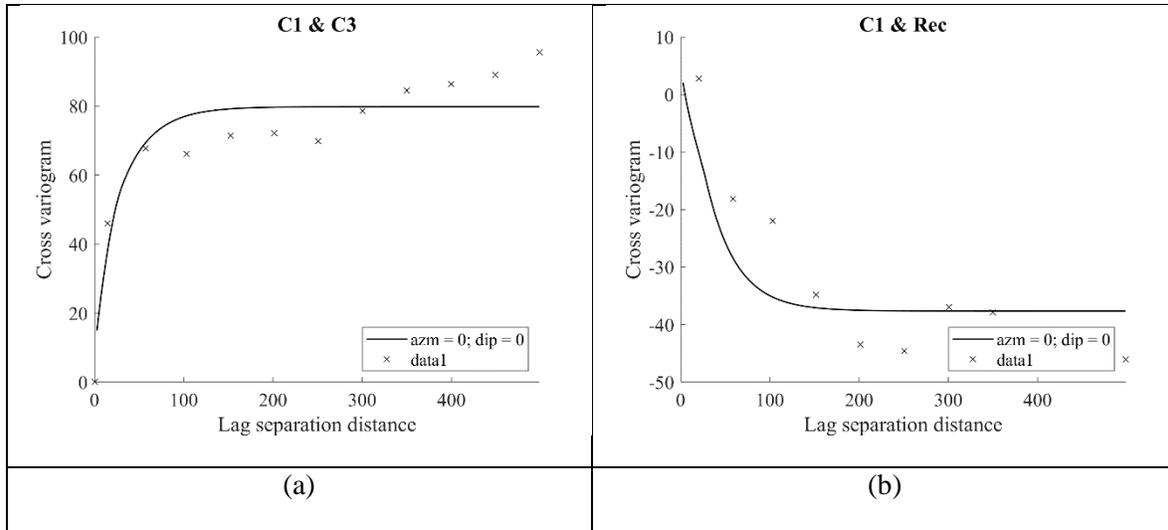


Figura 25 Variograma modelados (a) modelamiento de componente principal (CP) uno y tres; (b) modelamiento de vario grama de CP1 y Recuperación de cobre.

5.2.2 Revisión de la estimación de recuperación de cobre y componentes principales

Los resultados de este trabajo se pueden interpretar a través de dos propuestas que serán evaluadas en paralelo:

- 1) **Predicción de %RecCu directamente por cokriging** entre %RecCu (variable primaria) y componentes de espectros (variables secundarias).
- 2) **Predicción de %RecCu usando estimación de C1 al C4 pero interpretando con PLSR** utilizando la estimación de componentes que son transformadas a una estructura espectral, es decir, componentes transformados a un vector de reflectancia según longitud de onda similar a la proporcionada por el espectrómetro. Luego utilizando una regresión de mínimos cuadrados se obtiene %RecCu como resultado.

La primera propuesta puede ser considerada un método directo, siendo el resultado de la interpolación usado directamente como %RecCu, mientras que el segundo puede ser considerado indirecto porque requiere un procesamiento adicional de transformación de componentes principales a reflectancia para ser interpretado como %RecCu a través de PLSR.

Los resultados del método directo e indirecto son similares en términos de recuperación de cobre. Sin embargo, existen diferencias en la desviación estándar entre los métodos directo e indirecto, que pueden ser consideradas significativas porque se reduce a la mitad cuando se compara con el método directo.

La Figura 26 muestra el histograma del %RecCu en ambos métodos. La Tabla 4 cuantifica estas diferencias a través de estadísticas básicas.

Un leve sesgo entre el método directo e indirecto fue identificado, sin embargo, puede ser considerado poco significativo porque la mayor parte de los datos está situada dentro del rango de 85 y 95 %RecCu y ambas medias son similares. La Figura 27 muestra un gráfico de dispersión para ilustrar este sesgo.

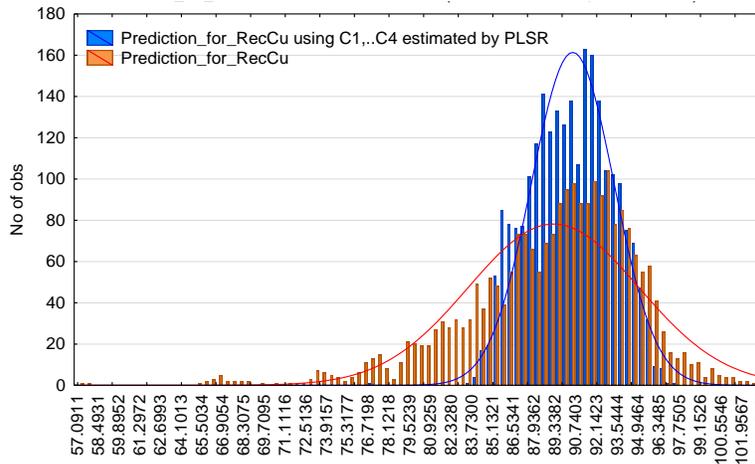


Figura 26 Muestra un histograma de %RecCu de las estrategias de estimación. Una manera directa y otra manera indirecta.

Tabla 4 Estadísticas básicas de resultados de %RecCu estimada

	número	Media	Minimo	Maximo	Std. Dev	C.V.
Rec Cu usando CoKriging	2404	89.07	57.09	103.83	5.73	15.5
Rec Cu usando C1,...C2, + PLSR	2404	90.36	76.79	97.73	2.77	32.6

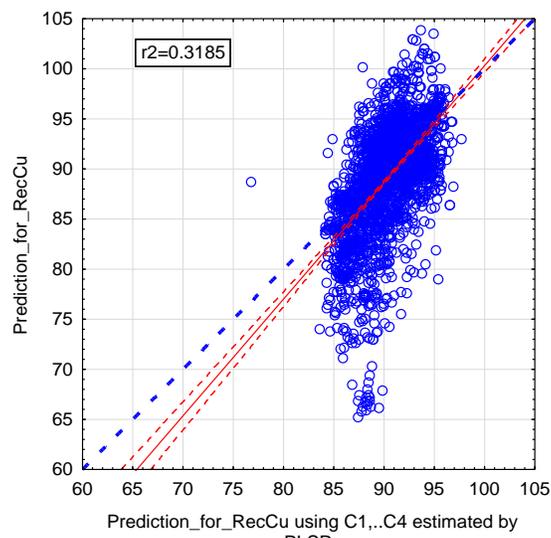


Figura 27 Nube de dispersión entre los métodos de estimación directa e indirecta.

5.2.3 Validación cruzada

Una revisión de los resultados de la validación cruzada, fue realizada para las siguientes alternativas propuestas:

- 1) Predicción directa de %RecCu usando cokriging entre %RecCu y componentes C1 al C4 de los espectros.
- 2) Predicción indirecta de %RecCu usando el cokriging de los componentes C1 a C4 de los espectros e interpretados con regresión de mínimo cuadrados parciales.

Los resultados refrendan las diferencias observadas en 5.2.2. que se pueden resumir en:

Tabla 5 Estadísticas básicas de los resultados de la validación cruzada

	número	Media	Minimo	Maximo	Std. Dev	C.V.
Recuperación de cobre ensayada	176	90.4	55.64	98.6	6.47	14.0
Rec Cu usando cokriging	176	89.89	68.48	101.17	5.87	15.3
Rec Cu usando C1,...C2, + PLSR	176	91.91	81.94	99.82	3.94	23.3

El promedio de %RecCu de los métodos directo e indirecto mostró diferencias poco significativas cuando se compara con el promedio de los datos reales ensayados. La Tabla 5 muestra que éstas no son mayores a un punto y medio de recuperación.

El método indirecto tiene menor desviación estándar que el método directo. Sin embargo, ambos fracasan al reproducir la desviación estándar de los datos reales.

Una revisión de algunos registros permitirá explicar el mal desempeño de ambos métodos directo e indirecto. La Figura 28 corresponde a un gráfico de probabilidad normal de %RecCu para los dos métodos y los valores reales ensayados.

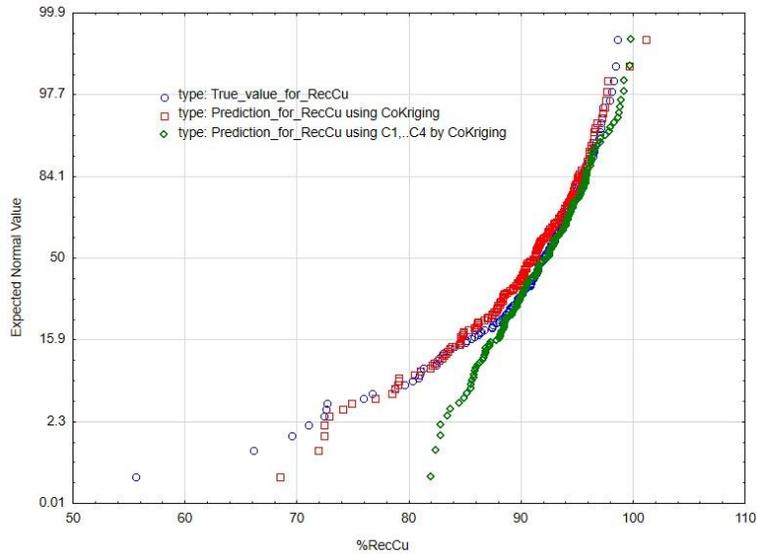


Figura 28 Gráficos de probabilidad normal de los resultados de validación cruzada. Valore real ensayado (azul), método directo (rojo), y método indirecto (verde).

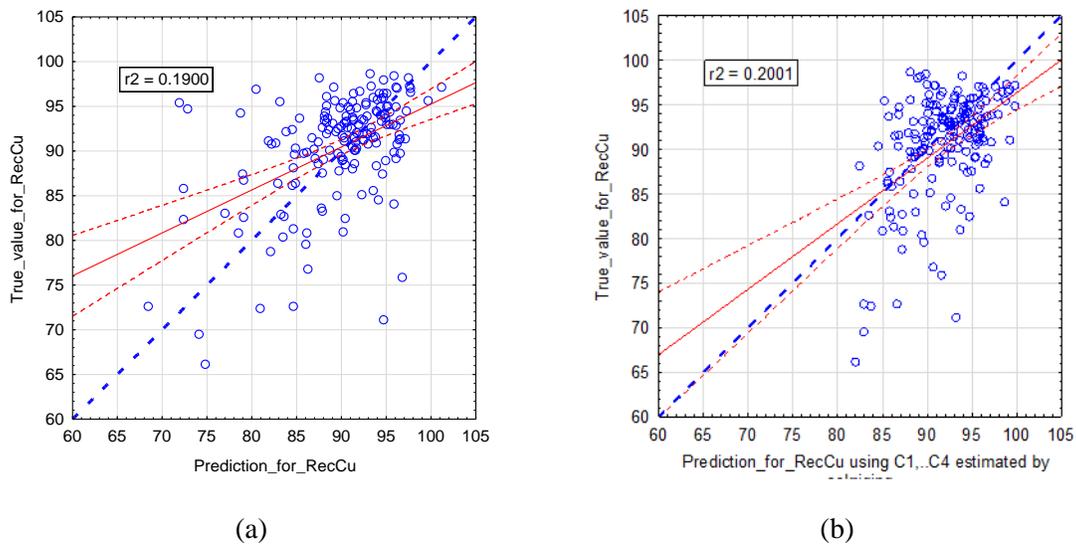
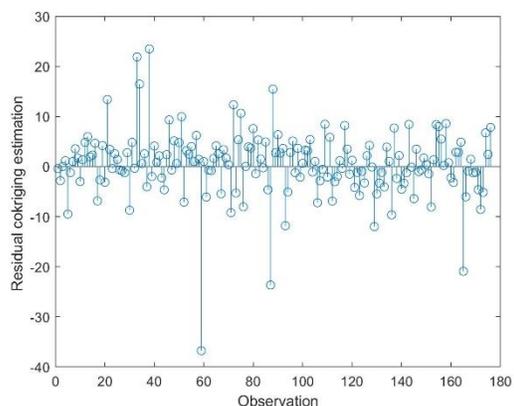
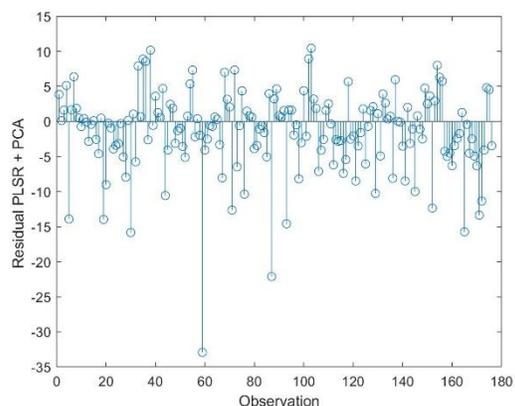


Figura 29 Muestra gráficos de dispersión de la estimación de %RecCu. (a) Método directo; (b) método indirecto

Una revisión de los errores residuales entre valor real y estimado fue realizada para ambos métodos propuestos de procesamiento de los resultados. Los siguientes registros muestran resultados de sobre estimación del %RecCu para cokriging, registros #59, #87, #165, #129 y #93, mientras que los registros #59, #87, #30, #165 y #93 para método usando componentes principales y procesados con PLSR. La Figura 30 muestra de manera gráfica los errores residuales, el eje de las abscisas muestra el orden de los registros, es decir, la numeración de los espectros. Ambos métodos presentaron problema para la correcta estimación de la recuperación de cobre.



(a)

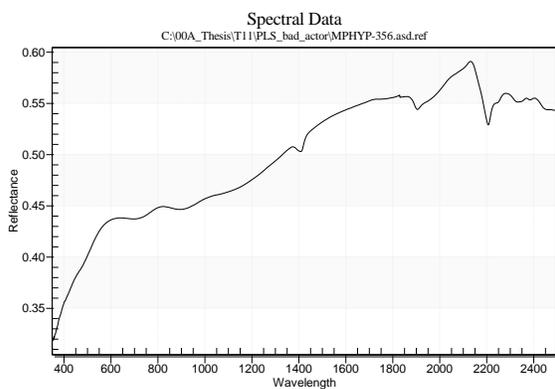


(b)

Figura 30 Errores residuales obtenidos para la validación cruzada

Una revisión detallada de los espectros fue realizada en el software Specmin®. Esta herramienta posee una librería especializada para espectros de pórfitos cupríferos. Tres espectros fueron revisados en #59(MPHYP-356.asd), #93(MPHYP-550.asd) y #165(MPHYP691.asd). La Figura 31 muestra una interpretación de los espectros con problema de predicción y, para ello se interpretó con una librería los minerales mayoritarios.

La presencia de minerales de arcillas tiene efectos negativos en los parámetros de recuperación metalúrgica. Una particularidad en estos ensayos con mal desempeño (#59, #93 y #165) fue la presencia de montmorillonita y caolinita, que fueron identificados a través espectrometría.



MPHYP-356 → %RecCu :55.6

Montmorillonita (49.80%)

Formula Química: $(\text{Na,Ca})_{0.3}(\text{Al,Mg})_2\text{Si}_4\text{O}_{10}(\text{OH})_2 \cdot n(\text{H}_2\text{O})$

Clase: filosilicato

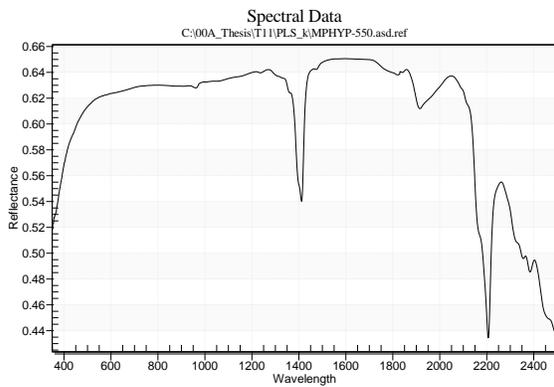
Grupo: esmectita

Ilita (50.20%)

Formula química: $\text{K}_{1-1.5}\text{Al}_4(\text{Si}_{6.5-7}\text{Al}_{1-1.5})\text{O}_{20}(\text{OH})_4$

Clase: filosilicato

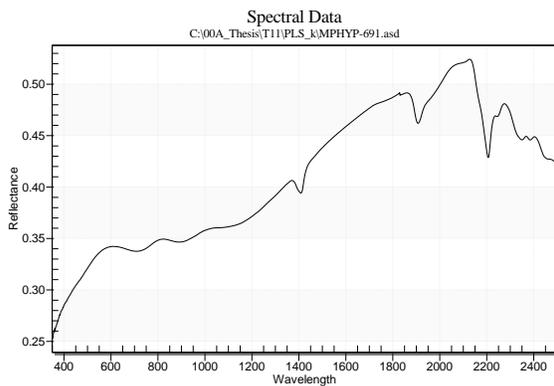
Grupo: ilita



MPHY-550 → %RecCu : 84.1

Montmorillonita (32.11%)
 Formula química: $(\text{Na,Ca})_{0,3}(\text{Al,Mg})_2\text{Si}_4\text{O}_{10}(\text{OH})_2 \cdot n(\text{H}_2\text{O})$
 Clase: filosilicato
 Grupo: esmectita

Kaolinite (67.89%)
 Formula química: $\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$
 Clase: filosilicato
 Grupo: caolinita



MPHY-691 → %RecCu : 75.9

Illita (46.54%)
 Formula química: $\text{K}_{1-1,5}\text{Al}_4(\text{Si}_{6,5-7}\text{Al}_{1-1,5})\text{O}_{20}(\text{OH})_4$
 Clase: filosilicato
 Grupo: illita

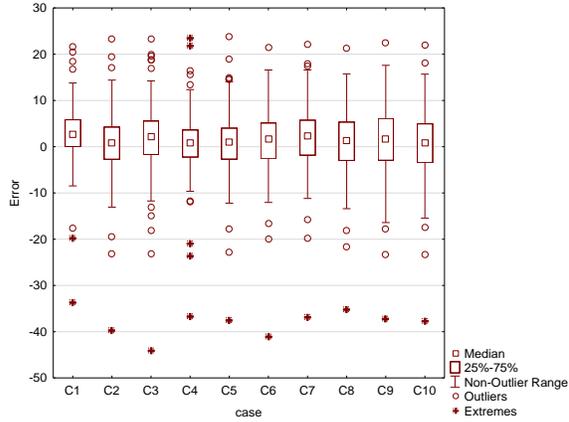
Kaolinite (53.46%)
 Formula química: $\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$
 Clase: filosilicato
 Grupo: caolinita

Figura 31 Información de especies minerales mayoritarias en espectros con mal desempeño.

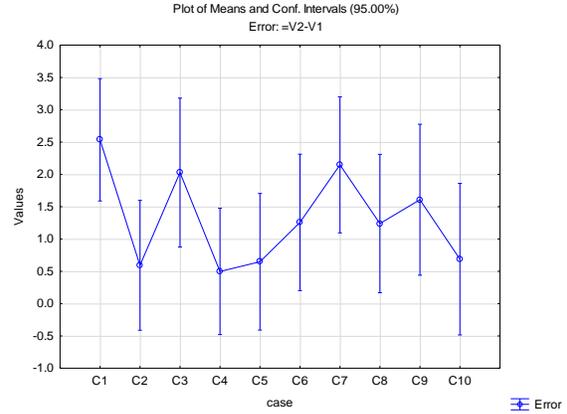
5.2.4 Revisión del número óptimo de componentes

El número de cuatro componentes principales utilizadas para la estimación de %RecCu se basó en el porcentaje de variabilidad. La Figura 32 muestra un ejercicio para revisar el error de predicción sujeto al número de componentes principales utilizados en la estimación.

A pesar que todos los errores fluctúan cerca de cero, la estimación con cuatro componentes tiene el menor error con media de 0.49 pero también con la menor dispersión de errores medidos con desviación estándar de 6.5.



(a)



(b)

Figura 32 (a) Box-plot de errores asociados a la validación cruzada en función del número de componentes. (b) Gráfico de media según el número de componentes.

Tabla 6 Estadísticas básicas de errores asociados a la validación cruzada en función de los 176 ensayos de flotación

Número de componentes	número	Media	Mínimo	Máximo	Std. Dev
1 componentes utilizado	176	2.53	-33.6	21.6	6.36
2 componentes utilizado	176	0.59	39.6	23.2	6.77
3 componentes utilizado	176	2.02	-44.2	23.2	7.74
4 componentes utilizado	176	0.49	-36.7	23.4	6.57
5 componentes utilizado	176	0.64	-14.5	23.7	7.1
6 componentes utilizado	176	1.25	-41.1	21.4	7.09
7 componentes utilizado	176	2.14	-36.9	22.1	7.07
8 componentes utilizado	176	1.23	-35.2	21.3	7.19
9 componentes utilizado	176	1.6	-37.1	22.4	7.85
10 componentes utilizado	176	0.68	-37.6	22	7.88

6 DISCUSIONES GENERALES

La principal propuesta de este estudio, sugirió la utilización de datos de espectrómetros para complementar la información de recuperación de cobre.

La utilización de la reflectancia como un proxy para la interpretación de minerales de ganga no es desconocida y está bien documentado el impacto de las arcillas en la cinética de flotación.

Durante la calibración pudieron ser revisados casos de presencia de arcillas.

La metodología propuesta permite la **reducción de la incertidumbre** con la incorporación de espectros a la vecindad espacial de los datos de flotación ensayado.

El método de interpolación utilizado es cokriging, mientras que el algoritmo de regresión utilizado es PLSR. Durante el desarrollo de este estudio se presentaron dos formas de utilizar los resultados:

- 1) Predicción de recuperación de cobre directamente usando cokriging entre %RecCu y componentes de espectros.
- 2) Predicción de recuperación de cobre usando estimación de C1 al C4, pero interpretando con PLSR utilizando la estimación de componentes que son transformadas a una estructura espectral, es decir, componentes transformadas a un vector de reflectancia y luego usando regresión de mínimos cuadrados, de lo que se obtiene %RecCu como resultado final.

Los datos finales obtenidos para la etapa de validación cruzada y predicción muestran que ambos métodos se diferencian por una disminución en la desviación estándar de los resultados.

El método indirecto tiene una menor desviación estándar de sus predicciones y no se observan problemas de estimación sobre 100%RecCu como los observados con el método directo.

La metodología propuesta como método indirecto muestra una innovadora aplicación de utilización de datos espectrales para un modelo de bloques. Tradicionalmente, los datos espectrales eran interpretados como contenido de arcillas y luego estos interpolados siguiendo los criterios de estimación, sin embargo, la solución tradicional únicamente entrega un número limitado de arcillas, es decir, provee una información parcial o incompleta. Lo innovador de la solución propuesta en el método indirecto, es proveer toda la información espectral en un modelo 3D, por lo tanto, facilita la **interpretación de arcillas** aplicando las técnicas tradicionales de interpretación espectral directamente sobre el modelo de bloques.

La

Figura 33 muestra una comparación de un espectro estimado y otro medido por el espectrómetro, siendo el estimado generado a partir de las cuatro componentes interpoladas con cokriging.

La utilización de datos espectrales tiene como propuesta de valor complementar los resultados de flotación de cobre con información de la mineralogía de ganga. Sin embargo, los datos obtenidos según la Figura 31 mostraron que el modelo frente a la presencia de arcillas no castigó el resultado de la recuperación de cobre. Ambos métodos erraron en su predicción, por lo tanto, la discusión

debería centrarse en los **criterios de aceptabilidad para una calibración** razonable para su utilización en el ámbito minero.

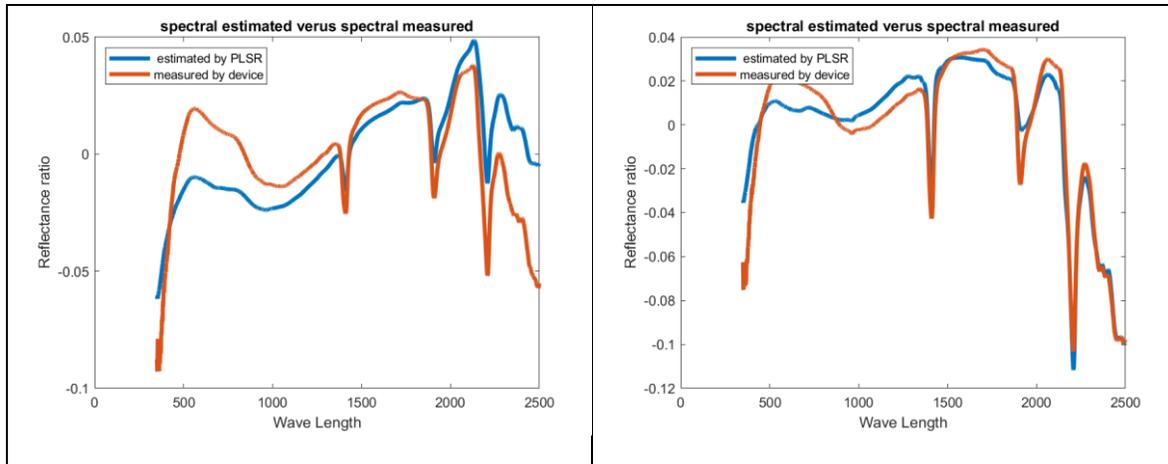


Figura 33 Muestra los resultados de la validación cruzada entre espectro medido y estimación de cokriging

La utilización del rango completo de la longitud de onda debe ser revisada porque las arcillas de importancia se registran en intervalos definidos. Algunos autores (Johnson, Curtis & Browning, David & Pendock, Neil, 2019) desarrollaron un trabajo similar utilizando 41 intervalos, evitando el ingreso de ruido.

Este estudio no evaluó alternativas para restringir o subdividir dominios para diferenciar poblaciones de arcillas, es decir, **etiquetar según los ambientes de alteración**.

La reflectancia según longitud de onda es relevante para cada mineral. Para ilustrar esto, los minerales ricos en hierro tienen menor reflectancia que los minerales de arcillas en el rango de 350nm a 1000nm.

La

Figura 34 muestra dos minerales de la clase filosilicatos, el primero llamado biotita ($\text{BiMA1} - \text{K}(\text{Mg,Fe})_3(\text{AlSi}_3\text{O}_{10}(\text{OH})_2)$) y el segundo llamado illita ($\text{IIIAZ2f} - \text{K}_{1-1.5}\text{Al}_4(\text{Si}_{6.5-7}\text{Al}_{1-1.5})\text{O}_{20}(\text{OH})_4$), ambos pertenecientes al grupo de las micas.

La imagen es útil para ilustrar que, mientras la reflectancia de la biotita aumenta según aumento de la longitud de onda, ésta es inversamente proporcional para la illita.

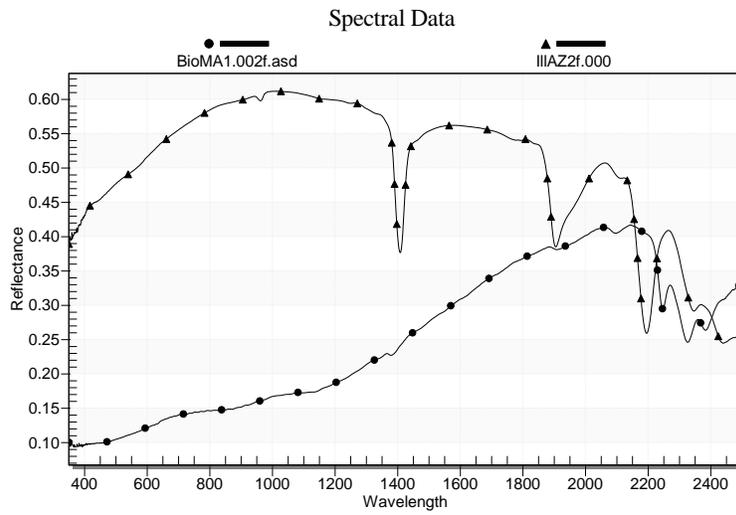


Figura 34 Muestra dos minerales de ganga Biotita(BioMA1) e Ilita(IIAZ2f) de librería Specmin®Pro

7 CONCLUSIONES

Lo expuesto a lo largo de este estudio permite llegar a las siguientes conclusiones:

La incertidumbre debido a la ausencia de datos de ensayos de flotación de cobre puede ser reducida con la utilización de datos espectrales, reduciendo esta última como componentes principales para su interpolación como una base de datos complementaria a los ensayos de flotación.

La revisión de los algoritmos de regresión PLSR y PCR muestra que la utilización de datos de espectrómetros debe ser manejada con regresión de mínimos cuadrados parciales, porque esta técnica es capaz de relacionar la *variable respuesta* (en este caso, %RecCu) y la *variable explicativa* proveniente de la información espectral a través de la covarianza de sus variables latentes, es decir, la convergencia de la solución relaciona la descomposición de la variable respuesta y la variable explicativa.

La revisión de alternativas de predicción de un modelo 3D de recuperación de cobre usando cokriging, generó un modelo entre %RecCu y cuatro componentes principales asociado a una reducción de dimensionalidad de la información de los espectros.

Dos opciones fueron revisadas: Predicción de %RecCu directa por cokriging (método directo) y predicción de %RecCu usando estimación de C1 al C4 pero interpretando con PLSR (método indirecto).

Las medias de la predicción de %RecCu de ambas alternativas tienen diferencias poco significativas inferiores a 1.4%, sin embargo, el método indirecto muestra dos características:

- a) **Primero:** Las estimaciones de % RecCu muestra una menor desviación estándar de sus resultados.
- b) **Segundo:** Las estimaciones de %RecCu del método indirecto son consistentes con el umbral de recuperación inferiores a 100%, situación que no ocurre con el método directo porque estimó bloques sobre dicho umbral.

La etapa de validación cruzada trabajó en tres escenarios: (a) predicción % Rec Cu usando cokriging; (b) predicción %RecCu usando la estimación de C1 a C4, para luego convertir a un espectro y procesar con PLSR; (c) predicción de %RecCu utilizando su espectro procesado con PLSR. Los resultados muestran un desempeño levemente mejor de la alternativa (b), pero se observa una variabilidad de las predicciones en la alternativa (a).

Una revisión de los errores de predicción de %RecCu debido al número componentes usado fue realizado donde se evaluaron 10 escenarios. El escenario con cuatro componentes principales muestra ser aquel con menor error residual entre valores reales y estimados. Además, este escenario de cuatro componentes presenta una menor desviación estándar.

En conclusión, la alternativa de incorporar datos de espectrómetros fue implementada y muestra una alternativa para reducir la incertidumbre en lugares sin información, siendo ésta una información

complementaria de bajo costo de implementación. La alternativa sugerida en este trabajo es la estrategia de interpolación de las componentes principales para luego procesar con una técnica de regresión.

La calidad de la información espectral debe ser revisada, si bien, la eliminación de outliers fue descrita como parte del proceso, ésta no fue desarrollada en su totalidad para evaluar impactos negativos en la predicción.

BIBLIOGRAFÍA

- Biernacki, C., & Govaert, G. (2007). Choosing models in model based clustering and discriminant analysis. *Journal of Statistical Computation and Simulation*, vol. 1(Issue 64), 49-71.
- Bioucas-Dias José et al. (2012). *Hyperspectral Unmixing Overview: Geometrical*. Journal of selected topics in applied earth observations and remote sensing, vol. 5, no. 2, april 2012.
- Brook, N. (april de 1985). The equivalent core diameter method of size and shape correction in point load testing. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics*, 22(2), 61-70.
- Caceres, A. (2010). *Simulación conjunta de unidades geológicas y leyes de cobre en el sector Sur-sur del deposito Rio Blanco - Los Bronces*. Universidad de Chile. Santiago: Memoria para optar al título de geólogo.
- Carmona, S. (2009). *Mejoramiento de la interpretación geológica mediante el uso de técnicas estadísticas multivariadas*. Universidad de Chile. Santiago: Tesis de grado para optar al grado de magister.
- Cornejo, J. (2009). *modelamiento geoestadístico de abundancias de minerales en el yacimiento Radomiro Tomic*. Universidad de Chile. Santiago: Memoria para optar al título de ingeniero civil de minas.
- De la Quintana, A. (2008). *Simulación de Unidades Litológicas en el Yacimiento Mansa Mina*. Universidad de Chile. Santiago: Memoria para optar al título de ingeniero civil de minas.
- Dirección de Modelamiento Geometalúrgico – Codelco Norte. (2011). *Aplicación de herramientas de inteligencia de negocios en modelamiento geometalúrgico*. Antofagasta: Mine Planning 2011.
- Donald Burns; Emil Ciurczak. (2008). *Handbook of Near - Infrared Analysis* (Vols. ISBN 978-0-8493-7393-0). CRC Press Taylor & Francis Group.
- Egaña, M. (2008). *Geoestadística aplicada a parámetros geotécnicos*. Universidad de Chile. Santiago: Memoria para optar al título de ingeniero civil de minas.
- Emery, X. (2000). *Geoestadística lineal* (Primera ed.). Santiago, Chile: Departamento de Ingeniería de Minas de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.
- Emery, X., & Cornejo, J. (2010). Truncated Gaussian simulation of discrete-valued ordinal coregionalized variables. *Computers & Geosciences*, Vol. 36, 1325–1338.

- Emery, X., & Silva, D. (June de 2009). Conditional co-simulation of continuous and categorical variables for geostatistical applications. *Computers & Geosciences*, Volume 35(Issue 6), 1234 -1246.
- Ferraudo, Antonio S. (2010). *Técnica de analisis multivariada*. Universidad Estatal Paulista UNESP: StatSoft South America.
- Geladi, P., & Kowalski, B. (1986). Partial Least-Squares Regression: A tutorial. *Elsevier Science Publishers*, 1-17.
- Goetz, A.F.H., et al. (2008). *Rapid Gangue Mineral concentration measured over conveyors by NIR reflectance spectroscopy*. miner engeenier.
- Guardiano, Srivastava. (1993). Multivariable Geostatistics Beyond bivariate Moments. In Soares, *Geostatistics Troia* (pp. 133-144). Kluwer, Dordrecht, V.1.
- Hartigan, J. A. (1972). Direct Clustering of a Data Matrix. *Journal of the American Statistical Association*, VOL.67(No.337), 123-129.
- Hubert, M. (2002). *A robust PCR method for high-dimensional regressors*. Belgium .
- Hurtado, S. (2009). *simulación de variables categóricas considerando estadísticas de patrones*. Universidad de Chile. Santiago: Memoria para optar al título de ingeniero civil de minas.
- Johnson, Curtis & Browning, David & Pendock, Neil. (2019). *Hyperspectral Imaging Applications to Geometallurgy: Utilizing Blast Hole Mineralogy to Predict Au-Cu Recovery and Throughput at the Phoenix Mine, Nevada*. Economic Geology. 10.5382/econgeo.4684.
- Mahalanobis. (1936). *On the generalized distance in statistics*. calcutta: Proceedings of the National Institute of Sciences (Calcutta) 2:49-55.
- Ortiz, D. (2017). *Estudio exploratorio del efecto de la caolinita en el comportamiento electroquímico de la calcopirita en el proceso de flotación*. Memoria de Título para optar al Título de Ingeniero Civil Metalúrgico.
- Robinson, Iain; MacArthur, Chris. (2015). *Field Spectroscopy Facility Post Processing Toolbox*. MATLAB Central File Exchange. Obtenido de (<https://www.mathworks.com/matlabcentral/fileexchange/31547-field-spectroscopy-facility-post-processing-toolbox>)
- Shlens, j. (2003). *A tutorial on Principal Component Analysis*. not published -(jonshlens@ucsd.edu).
- Smith, I. (2002). *A tutorial on Principal Components Analysis*. Obtenido de <http://www.cs.montana.edu/~harkin/courses/cs530/scilab/macros/cov.sci>
- Strebelle. (2009). *Multiple Point Statistics (MPS) Simulation with Enhanced Computational Efficiency*. Chevron U.S.A. Inc, San Ramon, CA: United State Patent N° US 1,516,055 B2.

- Strebelle, S. (2002). Conditional Simulation of Complex Geological Structures using Multiple - Point Statistics. *Mathematical Geology*, Vol. 34. N° 1 pp 1-21.
- Thuro, K; Plinninger R.J. (2001). Scale effects in rock Strength properties. *Rock Mechanics - a challenge for society*,, 169-180.
- Velde, B. & Meunier, Alain. (2008). *The Origin of Clay Minerals in Soils and Weathered Rocks*. 0.1007/978-3-540-75634-7.
- Velde, B. (1977). *Clays and Clay Minerals in Natural and Synthetic Systems*. Elsevier ISBN: 978-0-444-41505-9.
- Vilches, f. (2018). *Influencia de arcillas en la flotación de material calcopiritico*. memoria para optar al título de ingeniero civil metalúrgico, universidad técnica federico santa maria.
- Wackernagel. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Berlin: Springer.
- Williams. (2016). *Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions*. Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 36, No. 2.
- Wills. (2006). Mineral Processing Technology. *Elsevier*, 267.
- Witten, I., & Eibe, F. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. United States of America: Elsevier.
- Wold, H. (1975). *Quantitative Sociology - international perspectives on mathematical and Statistical Modelling*. New York: Academic Press inc.
- Y. Wang, Y. Peng, T. Nicholson, and R. Andreas. (2015). Applied Clay Science The different effects of bentonite and kaolini on copper flotation. *appl clay Science*, Vol 114 pp48-52.
- Zhang, M., & Peng, Y. (2015). Effect of clay minerals on pulp rheology and the flotation of copper and gold minerals. *Minerals Engineering*, vol. 70, pp. 8–13.