



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA DE MINAS

IMPLEMENTACIÓN DE METODOLOGÍA PARA DETERMINAR DOMINIOS
GEOMETALÚRGICOS DE ESTIMACIÓN

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL DE MINAS

DAVID ALONSO ROSALES FERNÁNDEZ

PROFESOR GUÍA:
JULIÁN ORTIZ CABRERA

MIEMBROS DE LA COMISIÓN:
XAVIER EMERY
ALEJANDRO CÁCERES SAAVEDRA

SANTIAGO DE CHILE
2014

Resumen

En el negocio minero es un asunto de vital importancia la evaluación de la rentabilidad esperada, que permita lograr un desempeño operacional óptimo con la menor incertidumbre posible. Actualmente, los modelos de estimación no consideran en mayor medida las variables de respuesta metalúrgica asociadas a las características del yacimiento, lo que conlleva a una evaluación y planificación no óptima, tanto minera como en planta, afectando en cierta medida una correcta proyección de los retornos económicos esperados.

Este trabajo de título tiene como objetivo principal proponer e implementar una metodología que permita definir, de forma no asistida o con poca intervención del usuario, dominios geometalúrgicos a partir de una base de datos real. Los datos utilizados en esta memoria fueron obtenidos de Mina Escondida, los cuales contienen información de geoquímica, asociaciones geológicas (litologías, alteraciones y mineralización) y recuperación de cobre por flotación rougher. Entre los objetivos específicos, se destaca realizar un modelo predictivo de recuperación de cobre por cada dominio definido, además de desarrollar un análisis de sensibilidad con respecto a la capacidad de los algoritmos de agrupamiento de clasificar los datos.

La metodología consiste en primer lugar en la construcción de una base de datos unificada a partir de los datos disponibles. Luego se realiza un estudio exploratorio de datos para después iniciar la minería de datos, consistente en la selección de variables, análisis de componentes principales y clustering (k-Means y Ward), que definirán los dominios. Una vez realizada esta labor se determina un modelo predictivo multilíneal de la recuperación de cobre por cada dominio obtenido, para finalmente efectuar un análisis de sensibilidad con respecto a la cantidad de clusters definidos y su capacidad de clasificar los datos.

Una vez obtenidos y analizados los dominios geometalúrgicos, se pudo concluir que esta metodología es útil ya que permite determinar de forma no supervisada estos clusters. Sin embargo, igualmente se requiere de la interpretación del investigador a cargo, ya que dependiendo del número de clusters definidos, se pueden generar unidades geometalúrgicas redundantes entre sí que no mejoren la calidad de la clasificación. En cuanto a los modelos predictivos, si bien una regresión multilíneal es de fácil aplicación e interpretación, ésta no ajusta de gran manera al presentar un coeficiente de determinación relativamente bajo, por lo que se recomienda ver alternativas tales como redes neuronales o árboles de decisión.

Abstract

In the mining business is a matter of vital importance the assessment of its likely profitability, enabling optimal operational performance with the lowest possible uncertainty. The current estimation models do not consider metallurgical response variables related to the characteristics of the reservoir, which leads to a non-optimal evaluation and planning in both mining and processing plant, affecting in some extent a correct projection of the expected economical return.

This thesis has a main objective, which is to propose and implement a methodology to define geometallurgical domains from a database with little user intervention or none at all. The database utilized was obtained from Mina Escondida, which contains information of geochemical variables, geological associations (lithology, alteration and mineralization) and copper recovery by rougher flotation. There are a few secondary objectives, the first one is to make a predictive model for copper recovery for each defined domain, in addition to a sensitivity analysis related to the ability of clustering algorithms to classify the data.

This methodology starts with building a unified database from the available data, then an exploratory study is made, after that data mining is performed, involving the selection of variables, principal component analysis and clustering (k-Means and Ward), which will define the domains. Once the domains are defined, a multivariable predictive model for copper recovery is obtained for each domain, and finally a sensitivity analysis is made to visualize the ability of geological classification depending on the number of clusters defined.

Once the geometallurgical domains are obtained and analyzed, the methodology proved to be useful because it allows the definition of domains in an unsupervised way, or with little supervision. However, it also requires the interpretation of the researcher in charge, because depending on the number of clusters, redundant metallurgic units could be created and may not improve the quality of classification. As for the predictive models, although a multilinear regression is easy to apply and to interpret, they have a low coefficient of determination, so it is recommended to seek alternatives such as neural networks or decision trees.

*The most difficult thing is the decision to act,
the rest is merely tenacity.*

-Amelia Earhart

Agradecimientos

Hoy que esta etapa plena de aprendizaje está a punto de culminar, deseo agradecer a todas las personas que de una u otra forma me acompañaron en este difícil, pero gratificante camino.

En primer lugar, quiero agradecer a mis padres por todo el apoyo y cariño incondicional que me han entregado durante toda mi vida y en especial en esta etapa universitaria. A mi madre por enseñarme a ser perseverante, resiliente y responsable en todo lo que hago. A mi padre por inculcarme el gusto por las matemáticas y la física. También quiero agradecer a mi hermana Camila y al resto de mi familia, ya que ellos también fueron parte de este crecimiento.

Gracias a todos mis amigos de primer año, los que me han apoyado en las buenas y en las malas: Rodrigo, Sebastián, Juan, Alonso, Werner, Marcela, Paula, Ignacio y Juano. Mis disculpas si he omitido a alguien, ustedes saben que también fueron parte de todo esto.

Mi especial agradecimiento a Daniela, mi polola, por acompañarme en esta etapa, apoyándome en mis frustraciones y alegrías, dándome todo su cariño cuando yo sentía que las cosas no estaban funcionando.

A todos mis compañeros mineros, con los que he compartido no solo en clases y trabajos, sino también en más de alguna actividad lúdica. Agradezco en especial a Giovanni, Antonio y Rubén.

Quiero también dedicarle un espacio a todos los integrantes de ALGES, agradecerles por la camaradería, los *mining nights* y por incluirme como si fuera un «computín» más.

A mis compañeros de oficina en el AMTC, con los que compartí todos estos meses de memoria. Específicamente a Pía y Efraín por hacerme clases de geología *for dummies*. A Roberto, Felipe, Ingrid, Oscar y Mauricio por lograr que los días de trabajo me fueran agradables.

A mi profesor guía Julián Ortiz por confiar en mis capacidades, por tener siempre una buena disposición a resolver mis dudas y a orientar mi trabajo. También agradezco a mi profesores de comisión, Xavier Emery y Alejandro Cáceres, por el apoyo y tiempo dedicado.

Muchas gracias a todos y cada uno. Apago las luces y cierro la puerta.

Tabla de contenido

1. Introducción	1
1.1. Motivación	2
1.2. Objetivos	3
1.2.1. General	3
1.2.2. Específicos	3
1.3. Alcances	3
2. Antecedentes	4
2.1. Pórfido cuprífero	4
2.1.1. Alteraciones hidrotermales	4
2.2. Marco geológico de mina Escondida	7
2.2.1. Litología	7
2.2.2. Alteración	9
2.2.3. Mineralización	11
2.3. Flotación de minerales de cobre	13
2.4. Análisis multivariable	14
2.4.1. Estudio exploratorio de datos	14
2.4.2. Coeficiente de correlación lineal de Pearson	16
2.4.3. Matriz de correlación	17
2.4.4. Análisis de componentes principales	17
2.4.5. Clustering	20
2.5. Modelo predictivo	23
2.5.1. Selección de variables	23
2.5.2. Regresión multilínea	24
2.5.3. Coeficiente de determinación	25
3. Metodología	27
3.1. Construcción base de datos	27
3.2. Estudio exploratorio de datos	27
3.3. Selección de variables	28
3.4. Análisis de componentes principales	28
3.5. Caso base	28
3.5.1. Clustering	29
3.5.2. Análisis de clusters	29
3.5.3. Modelo predictivo	29
3.6. Análisis de sensibilidad	29

4. Caso de estudio	31
4.1. Bases de datos originales	31
4.1.1. Geoquímica	31
4.1.2. Geología	32
4.1.3. Recuperación	32
4.2. Base de datos final	33
4.2.1. Variables sintéticas	34
4.2.2. Tratamiento de variables geológicas	35
4.3. Estudio exploratorio de datos	39
4.3.1. Datos de geoquímica	39
4.3.2. Variables sintéticas	41
4.3.3. Datos de recuperación	42
4.3.4. Datos de geología	44
5. Definición de dominios geometalúrgicos	47
5.1. Selección de variables	47
5.2. Análisis de componentes principales	49
5.3. Clustering	54
5.3.1. k-Means	56
5.3.2. Ward	61
5.3.3. Discusiones	66
5.4. Modelo predictivo	69
5.4.1. k-Means	70
5.4.2. Ward	72
5.4.3. Discusiones	75
5.5. Análisis de sensibilidad	76
5.5.1. 4 clusters	76
5.5.2. 5 clusters	79
5.5.3. 6 clusters	82
5.5.4. Discusiones	84
6. Conclusiones y recomendaciones	86
Bibliografía	88
A. Estudio exploratorio de datos	90
A.1. Base de datos original	90
A.2. Base de datos final	92
A.2.1. Matriz de correlación	92
B. Antecedentes	104
B.1. Clustering jerárquico	104
C. Enfoque tradicional de definición de dominios geometalúrgicos	105
C.1. Dominios	105
C.2. Variables seleccionadas	106
C.2.1. Dominio A	106
C.2.2. Dominio B	106

C.2.3. Dominio C	107
C.2.4. Dominio D	107
C.3. Estadísticas	108
C.3.1. Dominio A	108
C.3.2. Dominio B	109
C.3.3. Dominio C	110
C.3.4. Dominio D	111
C.4. Modelo predictivo	111
C.4.1. Dominio A	112
C.4.2. Dominio B	112
C.4.3. Dominio C	113
C.4.4. Dominio D	113
C.5. Conclusiones	113

Índice de tablas

2.1. Ejemplo de matriz de correlación	17
2.2. Ejemplo de matriz de confusión	22
4.1. Elementos presentes en la base de datos de geoquímica	32
4.2. Códigos litologías	36
4.3. Códigos alteraciones	37
4.4. Códigos mineralizaciones	37
4.5. Nuevos códigos de litologías	38
4.6. Nuevos códigos de alteraciones	38
4.7. Nuevos códigos de mineralizaciones	38
4.8. Coordenadas de los datos.	39
4.9. Estadísticas generales	40
4.9. Estadísticas generales	41
4.10. Estadísticas de las variables sintéticas	41
4.11. Estadísticas de recuperación	43
5.1. Correlaciones entre recuperación y variables geoquímicas y sintéticas. Ordenados de mayor a menor.	48
5.2. Componentes principales junto con sus varianzas explicadas y acumulada. Componentes 1 a 6	50
5.3. Componentes principales junto con sus varianzas explicadas y acumulada. Componentes 6 a 13	50
5.4. Estadísticas cluster A de k-Means	57
5.5. Estadísticas cluster B de k-Means	58
5.6. Estadísticas cluster C de k-Means	59
5.7. Estadísticas cluster D de k-Means	60
5.8. Estadísticas cluster A de Ward	62
5.9. Estadísticas cluster B de Ward	63
5.10. Estadísticas cluster C de Ward	64
5.11. Estadísticas cluster D de Ward	65
5.12. Matriz de confusión entre resultados de k-Means y Ward.	66
5.13. Modelo predictivo en Cluster A de k-Means	70
5.14. Modelo predictivo en Cluster B de k-Means	71
5.15. Modelo predictivo en Cluster C de k-Means	71
5.16. Modelo predictivo en Cluster D de k-Means	72
5.17. Modelo predictivo en Cluster A de Ward	73

5.18. Modelo predictivo en Cluster B de Ward	73
5.19. Modelo predictivo en Cluster C de Ward	74
5.20. Modelo predictivo en Cluster D de Ward	75
A.1. Estadísticas de leyes	90
A.1. Estadísticas de leyes	91
A.2. Estadísticas de recuperación	91
A.3. Estadísticas de variables sintéticas en base de datos original.	92
A.4. Matriz de correlación. Parte 1 de 7	92
A.4. Matriz de correlación. Parte 1 de 7	93
A.4. Matriz de correlación. Parte 1 de 7	94
A.5. Matriz de correlación. Parte 2 de 7	94
A.5. Matriz de correlación. Parte 2 de 7	95
A.6. Matriz de correlación. Parte 3 de 7	95
A.6. Matriz de correlación. Parte 3 de 7	96
A.6. Matriz de correlación. Parte 3 de 7	97
A.7. Matriz de correlación. Parte 4 de 7	97
A.7. Matriz de correlación. Parte 4 de 7	98
A.8. Matriz de correlación. Parte 5 de 7	98
A.8. Matriz de correlación. Parte 5 de 7	99
A.8. Matriz de correlación. Parte 5 de 7	100
A.9. Matriz de correlación. Parte 6 de 7	100
A.9. Matriz de correlación. Parte 6 de 7	101
A.10. Matriz de correlación. Parte 7 de 7	101
A.10. Matriz de correlación. Parte 7 de 7	102
A.10. Matriz de correlación. Parte 7 de 7	103
C.1. Variables seleccionadas en el dominio A. Metodología tradicional.	106
C.2. Variables seleccionadas en el dominio B. Metodología tradicional.	106
C.3. Variables seleccionadas en el dominio C. Metodología tradicional.	107
C.4. Variables seleccionadas en el dominio D. Metodología tradicional.	107
C.5. Estadísticas de recuperación. Dominio A de metodología clásica.	108
C.6. Estadísticas de recuperación. Dominio B de metodología clásica.	109
C.7. Estadísticas de recuperación. Dominio C de metodología clásica.	110
C.8. Estadísticas de recuperación. Dominio D de metodología clásica.	111
C.9. Resultados de regresión multilínea para dominio A, bajo diferentes parámetros.112	
C.10. Resultados de regresión multilínea para dominio B, bajo diferentes parámetros.112	
C.11. Resultados de regresión multilínea para dominio C, bajo diferentes parámetros.113	
C.12. Resultados de regresión multilínea para dominio D, bajo diferentes parámetros.113	

Índice de figuras

2.1. Zonas de alteración-mineralización generales en un pórfido cuprífero. Fuente: [20]	5
2.2. Asociaciones de minerales de alteración comunes en sistemas hidrotermales. Fuente: [4].	6
2.3. Fotografía satelital del Distrito Escondida. Fuente: [1]	7
2.4. Litologías presentes en mina Escondida. Modificado de [1]	9
2.5. Alteraciones presentes en mina Escondida. Modificado de [1]	10
2.6. Mineralizaciones presentes en mina Escondida. Modificado de [1]	13
2.7. Circuito convencional de flotación de cobre, en donde se destaca la etapa rougher, fuente de los datos experimentales de recuperación. Fuente: [12]	14
2.8. Ejemplo de histograma de frecuencias	15
2.9. Ejemplo de un diagrama de caja (boxplot)	16
2.10. Ejemplo del índice de correlación de Pearson. Fuente: [8]	17
2.11. Ejemplo de gráfico de sedimentación.	20
2.12. Ejemplo de agrupamiento <i>k-Means</i> . Fuente: [11]	21
2.13. Ejemplo de algoritmo LARS Lasso. En el eje de las abscisas está el coeficiente s y en eje de las ordenadas el coeficiente de la variable predictiva.	25
3.1. Esquema de la metodología a implementar	27
4.1. Cantidad de datos por cada tipo de variable geológica. (a): Litologías. (b): Alteraciones. (c): Mineralizaciones	36
4.2. Vista básica en planta de mina Escondida. En el cuadrado se encuentra la zona donde se distribuyen los datos a analizar. Modificado de [1]	39
4.3. Matriz de correlación con todas las variables presentes en el estudio, incluyendo variables sintéticas.	42
4.4. Histograma de recuperación de cobre.	43
4.5. Vista espacial de la recuperación de cobre. (a): Vista isométrica de sondajes con sus valores de recuperación de cobre. (b): Vista en el plano Este-Norte. (c): Vista en el plano Este-Cota. (d): Vista en el plano Norte-Cota.	44
4.6. Recuperación por tipo de litología	45
4.7. Recuperación por tipo de alteración	45
4.8. Recuperación por tipo de mineralización	46
5.1. Gráfico de sedimentación	51
5.2. Vista isométrica de las tres componentes principales.	51

5.3.	Visualización de resultados de PCA. (a): Vista bidimensional componentes 1 y 2. (b): Vista bidimensional componentes 1 y 3. (c): Vista bidimensional componentes 2 y 3.	52
5.4.	Presencia de arcillas en función de la primera componente principal	53
5.5.	Presencia de arcillas en función de la segunda componente principal	53
5.6.	Presencia de arcillas en función de la tercera componente principal	54
5.7.	Coficiente de silueta por cada algoritmo de clustering.	55
5.8.	Resultados de clustering (k-Means de 4 dominios) sobre datos de PCA. (a): Vista bidimensional componentes 1 y 2. (b): Vista bidimensional componentes 1 y 3. (c): Vista bidimensional componentes 2 y 3.	56
5.9.	Boxplots de recuperación por tipo de variable geológica en el cluster A definido por k-Means. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.	57
5.10.	Boxplots de recuperación por tipo de variable geológica en el cluster B definido por k-Means. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.	58
5.11.	Boxplots de recuperación por tipo de variable geológica en el cluster C definido por k-Means. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.	59
5.12.	Boxplots de recuperación por tipo de variable geológica en el cluster D definido por k-Means. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.	60
5.13.	Resultados de clustering (Ward de 4 dominios) sobre datos de PCA. (a): Vista bidimensional componentes 1 y 2. (b): Vista bidimensional componentes 1 y 3. (c): Vista bidimensional componentes 2 y 3.	61
5.14.	Boxplots de recuperación por tipo de variable geológica en el cluster A definido por Ward. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.	62
5.15.	Boxplots de recuperación por tipo de variable geológica en el cluster B definido por Ward. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.	63
5.16.	Boxplots de recuperación por tipo de variable geológica en el cluster C definido por Ward. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.	64
5.17.	Boxplots de recuperación por tipo de variable geológica en el cluster D definido por Ward. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.	65
5.18.	Boxplots de recuperación por cluster en cada método. (a): k-Means. (b): Ward.	67
5.19.	Proporciones de litologías para cada cluster definido por k-Means.	68
5.20.	Proporciones de alteraciones para cada cluster definido por k-Means.	68
5.21.	Proporciones de mineralizaciones para cada cluster definido por k-Means.	69
5.22.	4 clusters definidos por cada método (a) k-Means. (b) Ward	76
5.23.	Proporciones de litologías por tipo de método para 4 clusters. (a) k-Means. (b) Ward	77
5.24.	Proporciones de alteraciones por tipo de método para 4 clusters. (a) k-Means. (b) Ward	77

5.25. Proporciones de mineralizaciones por tipo de método para 4 clusters. (a) k-Means. (b) Ward	78
5.26. 5 clusters definidos por cada método (a) k-Means. (b) Ward	79
5.27. Proporciones de litologías por tipo de método para 5 clusters. (a) k-Means. (b) Ward	80
5.28. Proporciones de alteraciones por tipo de método para 5 clusters. (a) k-Means. (b) Ward	80
5.29. Proporciones de mineralizaciones por tipo de método para 5 clusters. (a) k-Means (b) Ward	81
5.30. 6 clusters definidos por cada método (a) k-Means. (b) Ward	82
5.31. Proporciones de litologías por tipo de método para 6 clusters. (a) k-Means. (b) Ward	83
5.32. Proporciones de alteraciones por tipo de método para 6 clusters. (a) k-Means. (b) Ward	83
5.33. Proporciones de mineralizaciones por tipo de método para 6 clusters. (a) k-Means. (b) Ward	84
A.1. Histograma de la recuperación de cobre a partir de los datos originales . . .	91
C.1. Boxplots de recuperación por tipo de variable geológica en el dominio A. Metodología original. (a): Recuperación por litologías. (b): Recuperación por mineralizaciones.	108
C.2. Boxplots de recuperación por tipo de variable geológica en el dominio B. Metodología original. (a): Recuperación por litologías. (b): Recuperación por mineralizaciones.	109
C.3. Boxplots de recuperación por tipo de variable geológica en el dominio C. Metodología original. (a): Recuperación por litologías. (b): Recuperación por mineralizaciones.	110
C.4. Boxplots de recuperación por tipo de variable geológica en el dominio D. Metodología original. (a): Recuperación por litologías. (b): Recuperación por mineralizaciones.	111

Capítulo 1

Introducción

Dentro de los múltiples objetivos de la evaluación de proyectos mineros, estimar la probable rentabilidad y lograr un efectivo desempeño operacional son tal vez los más importantes. Para lograr estos objetivos se debe contar con la mayor información posible, tanto en cuanto a las propiedades de la roca a explotar como de su posible desempeño en los procesos de recuperación del metal de interés. El conocimiento o desconocimiento de estas características impactarán en la economía del proyecto y su valor total.

Los modelos de recursos actuales consisten típicamente de tonelaje y leyes de mineral sobre una ley de corte, además de su distribución espacial. Datos como la dilución y recuperación metalúrgica son resultado de factores aplicados en base a los datos del modelo de recursos al evaluar la rentabilidad del proyecto. Considerando este contexto, la pregunta que uno debe hacerse es, ¿es esta estimación del valor de un proyecto, lo suficientemente fidedigna?, la premisa de este trabajo de memoria es que la incorporación de variables geometalúrgicas al modelo de recursos, tales como recuperación metalúrgica, consumo de ácido, consumo de reactivos y/o variables geomecánicas, entre otras, ayudan a un mejoramiento tanto en la predicción de retornos económicos como en la planificación de la mina y de planta de procesamiento del mineral.

El campo que estudia la incorporación de estos datos de respuesta metalúrgica en los modelos de recursos se denomina *geometalurgia*, la cual se define como una disciplina emergente que integra estas nuevas variables, creando modelos que relacionen propiedades intrínsecas de la roca con respuesta metalúrgica. Estos modelos geometalúrgicos están definidos en dominios, los cuales comparten características de respuesta metalúrgica bajo ciertos parámetros de roca, que no necesariamente coinciden en tamaño y en límites con los dominios definidos para las unidades geológicas.

Definir estos dominios geometalúrgicos es la base de este trabajo de título, labor para la cual se propuso una metodología no supervisada, que requiere poca intervención del investigador a cargo, sin embargo, estas herramientas ofrecen la flexibilidad para que el modelador imponga su conocimiento del fenómeno. Esta metodología utiliza herramientas de minería de datos, tales como selección de variables, análisis de componentes principales y clustering.

Este trabajo se enmarca dentro de Laboratorio ALGES, parte del Advanced Mining Technology Center (AMTC¹). Las herramientas aplicadas en este trabajo están implementadas en la librería de código abierto *scikit-learn* [16], desarrollada en lenguaje *Python*².

1.1. Motivación

La incorporación de datos de respuesta metalúrgica apuntan a una mejora en la calidad de información de los modelos de recursos. Este enriquecimiento de los datos implica múltiples beneficios no solo en la estimación de recursos, sino también en la optimización de la operación tanto en mina como en planta [5, 6]. La determinación de correctos dominios geometalúrgicos, en base a esta nueva información, implica una serie de mejoras en el proceso minero, entre las cuales se destacan:

- Predicción de elementos y/o minerales que pueden afectar el proceso de recuperación, como por ejemplo el contenido de arcillas [3].
- Mejora en la predicción de los retornos económicos, que es consecuencia de un secuenciamiento y/o planificación eficientes.
- Predicción de beneficios y costos.
- Optimización de los procesos de beneficio.
- Manejo eficiente de sustancias peligrosas, debido a la correcta predicción de la presencia de estos materiales, y
- Mejoras en la sustentabilidad de los proyectos, lo cual es de absoluta relevancia en una industria donde es cada vez más difícil obtener licencias operativas.

En resumen, la motivación principal de este trabajo de memoria es proponer una metodología que permita definir dominios geometalúrgicos y realizar modelos predictivos de variables geometalúrgicas con poca intervención de parte del investigador, con la finalidad de mejorar la calidad de información, ayudando no solo a definir un plan y diseño mineros, sino también a un correcto plan de procesamiento de minerales, con su consecuente mejora en los beneficios económicos y decisiones en planificación.

¹<http://www.amtc.cl/>

²<https://www.python.org/>

1.2. Objetivos

1.2.1. General

- Recomendar una metodología y las herramientas necesarias para definir dominios geometalúrgicos a partir de ensayos metalúrgicos e información de sondajes y geología.

1.2.2. Específicos

- Reducir la dimensionalidad de los datos, eliminando variables irrelevantes y/o redundantes.
- Desarrollar un modelo predictivo que relacione variables de geoquímica con recuperación de cobre por flotación rougher, para cada dominio definido.
- Realizar análisis de sensibilidad respecto a la cantidad de dominios calculados.
- Discutir la aplicación del modelo de definición de dominios geometalúrgicos.

1.3. Alcances

El estudio consiste en aplicar y estudiar los resultados de la metodología propuesta para la determinación de dominios geometalúrgicos. Esta metodología requiere poca intervención del investigador, salvo para ciertos análisis de sensibilidad que se deben realizar bajo ciertas condiciones. Este estudio será aplicado sobre una base de datos perteneciente al sector sur de Mina Escondida, la cual contiene información geoquímica, variables geológicas (litologías, alteraciones y mineralizaciones) y porcentaje de recuperación de cobre a través de flotación rougher.

Se espera que la metodología propuesta sea escalable a cualquier base de datos que incluya información tanto de leyes de elementos químicos, asociaciones geológicas y la(s) variable(s) geometalúrgica(s) a estudiar.

En este estudio no se dio mayor énfasis a los controles geológicos detrás de los dominios obtenidos ni a los modelos predictivos de recuperación de cobre. En resumen, se discute los resultados desde un punto de vista de la minería de datos aplicada y su capacidad de agrupar conjuntos de datos disímiles entre sí, además de medir la precisión de los modelos predictivos obtenidos.

Capítulo 2

Antecedentes

2.1. Pórfido cuprífero

Un pórfido cuprífero es un depósito mineral de baja ley de cobre y alto tonelaje. Se les llama pórfidos ya que generalmente se encuentran asociados a rocas ígneas intrusivas con fenocristales de feldespatos en una masa fundamental cristalina de grano fino.

Estos tipos de yacimientos están asociados a arcos magmáticos de márgenes continentales y a magmatismo calco-alcalino de composición intermedia, la mineralización ocurre de forma disseminada en vetillas, stockworks, brechas y como relleno [14, 20].

En la figura 2.1 se puede observar las alteraciones hidrotermales presentes en un pórfido cuprífero, las cuales se detallarán a continuación.

2.1.1. Alteraciones hidrotermales

Una alteración hidrotermal, en términos generales, es la respuesta mineralógica, textural y química de las rocas a un cambio ambiental generado tanto por fenómenos químicos y termales en presencia de agua caliente, vapor o gas. Las alteraciones hidrotermales ocurren a través de la transformación de fases minerales, crecimiento de minerales nuevos, disolución y/o precipitación de minerales junto con reacciones de intercambio iónico entre los minerales constituyentes de la roca y el fluido caliente que circuló por la misma [14]. Las alteraciones hidrotermales presentes en un pórfido cuprífero son:

2.1.1.1. Potásica

Zona ubicada en el núcleo del pórfido cuprífero, con una alta presencia de biotita y feldespato potásico junto con minerales accesorios como la actinolita, epidota, sericita y albita. En esta zona se produce mineralización de pirita-calcopirita, calcopirita \pm bornita, bornita

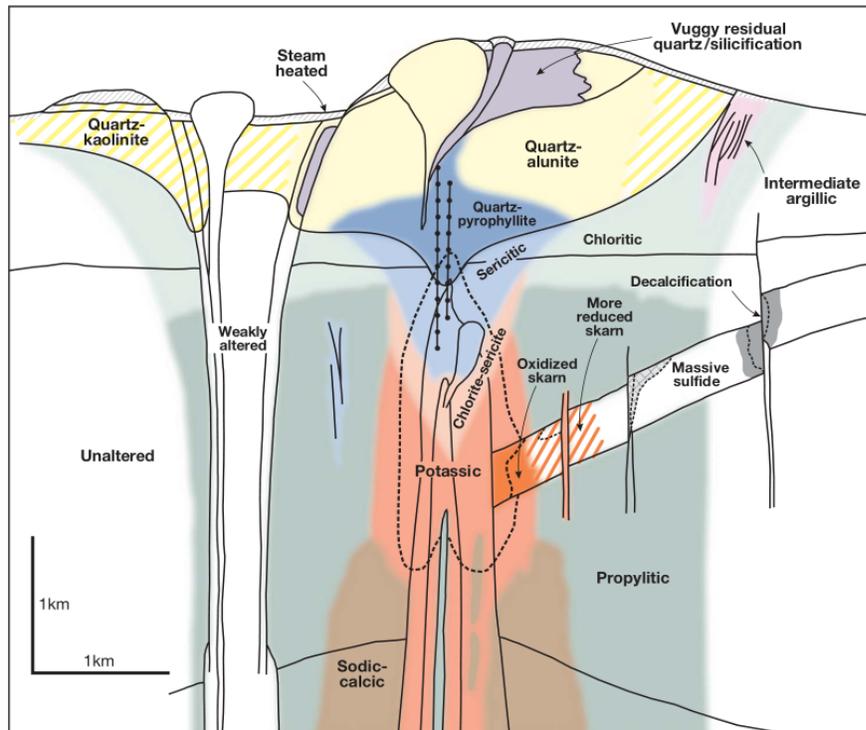


Figura 2.1: Zonas de alteración-mineralización generales en un pórfido cuprífero. Fuente: [20]

± digenita ± calcosina, las cuales son principales fuentes de mena o minerales de cobre de interés económico. Se origina en ambientes de pH neutro a alcalino, en altas temperaturas.

2.1.1.2. Propilitica

Zona asociada a la presencia de clorita y epidota, con o sin presencia de albita con minerales accesorios como la actinolita, hematita y magnetita. Se forma en condiciones de pH neutro a alcalino en un rango de temperatura entre 200 y 300 grados Celsius.

2.1.1.3. Fílica

Llamada también sericitica, esta zona se caracteriza por estar asociada principalmente a cuarzo y sericita, con minerales accesorios como clorita, illita y pirita. Se ubica espacialmente en la zona superior del pórfido cuprífero y se origina en un rango de pH entre 5 y 6 a una temperatura sobre los 250 grados Celsius.

2.1.1.4. Argílica intermedia

Se caracteriza por la presencia de arcillas (caolinita, smectita, montmorillonita o arcillas amorfas) y mayor o menor presencia de cuarzo. Ocurre en rangos de pH entre 4 y 5 y a temperaturas bajo los 300 grados Celsius.

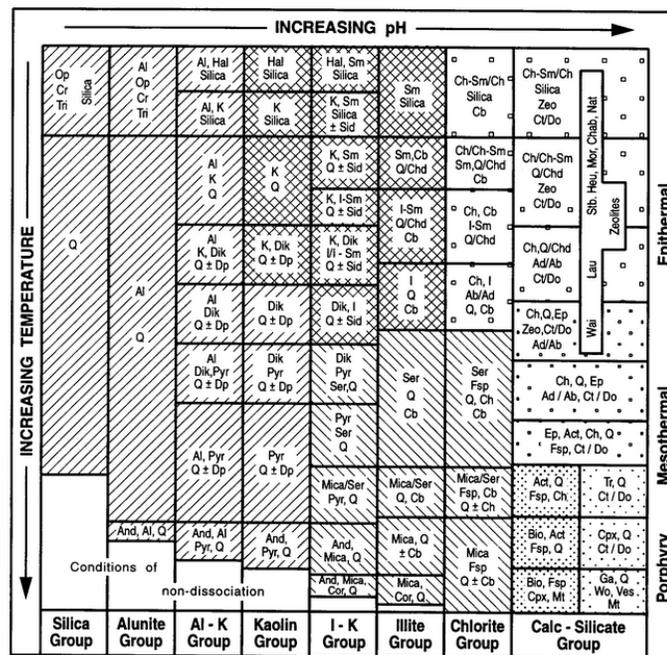
2.1.1.5. Argílica avanzada

Zona de fuerte alteración, con presencia de cuarzo oqueroso con o sin presencia de alunita, jarosita, caolín, pirofilita y piritita. Se genera en un amplio rango de temperaturas en un pH entre 1 a 3.5.

2.1.1.6. Calcosilicatada (*skarn*)

Corresponde a la zona donde ocurre transformación de rocas carbonatadas (calizas, dolomitas) a minerales calcosilicatados en zonas adyacentes a intrusivos. Se caracteriza por la presencia de granates (andradita y grosularita), wollastonita, epidota, diópsido, idocrasa, clorita y actinolita.

Corbett y Leach en 1998 [4] desarrollaron un diagrama de clasificación de tipo de alteración hidrotermal, en el que se incluyen los principales tipos de alteración ordenados en función del pH del fluido y de la temperatura (figura 2.2).



Mineral Abbreviations :

Ab - albite; Act - actinolite; Ad - adularia; Al - alunite; And - andalusite; Bio - biotite; Cb - carbonate (Ca, Mg, Mn, Fe); Ch - chlorite; Chab - chabazite; Chd - chalcedony; Ch-Sm - chlorite-smectite; Cor - corundum; Cpx - clinopyroxene; Cr - cristobalite; Ct - calcite; Do - dolomite; Dik - dickite; Dp - diasporite; Ep - epidote; Fsp - feldspar; Ga - garnet; Hal - halloysite; Heu - heulandite; I - illite; I-Sm - illite-smectite; K - kaolinite; Lau - laumontite; Mt - magnetite; Mor - mordenite; Nat - natrolite; Op - opaline silica; Pyr - pyrophyllite; Q - quartz; Ser - sericite; Sid - siderite; Sm - smectite; Stb - stilbite; Tr - tremolite; Tri - tridymite; Ves - vesuvianite; Wai - wairakite; Wo - wollastonite; Zeo - zeolite

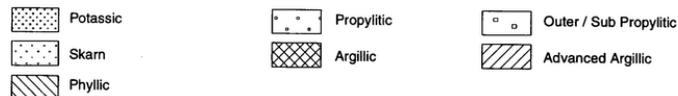


Figura 2.2: Asociaciones de minerales de alteración comunes en sistemas hidrotermales. Fuente: [4].

2.2. Marco geológico de mina Escondida

El yacimiento Escondida, junto con el yacimiento Escondida Norte, conforman el clúster Escondida, el cual geológicamente está ubicado dentro de la franja de los yacimientos tipo pórfido cuprífero del Eoceno Superior - Oligoceno Inferior, al que también pertenecen importantes yacimientos tales como Chuquicamata, Collahuasi, El Salvador, entre otros (figura 2.3).



Figura 2.3: Fotografía satelital del Distrito Escondida. Fuente: [1]

Escondida es un depósito porfídico de cobre con presencia adicional de molibdeno (Mo) y oro (Au) diseminado, caracterizado por su gran tonelaje y bajas leyes. En este sistema de pórfido existen varias fases de intrusión, en donde las más tempranas presentan leyes más altas y además existe mineralización en la roca caja [1].

2.2.1. Litología

Los yacimientos Escondida y Escondida Norte están asociados a un complejo intrusivo de composición monzonítica a granodiorítica, llamado *Complejo Intrusivo Feldespático Escondida* [22], el cual está en contacto con una intrusión de rocas volcánicas andesíticas

pertenecientes a la formación Augusta Victoria. Este complejo, a su vez, está intruido por cuerpos subvolcánicos tardíos de composición riolítica y por diques dacíticos.

2.2.1.1. Formación Augusta Victoria

Esta unidad está constituida esencialmente por rocas andesíticas, distribuidas principalmente en el SW del rajo Escondida y en el sector E de Escondida Norte. Corresponde a la roca caja del complejo subvolcánico mineralizador del yacimiento Escondida.

2.2.1.2. Complejo Intrusivo Feldespático Escondida (CIFE)

El Complejo Intrusivo Feldespático Escondida (CIFE) se define como un grupo de intrusivos que muestran una variación textural y composicional, pero que forman parte de un mismo sistema. Posee rocas de composición intermedia, variando de monzonitas a granodioritas con textura porfídicas. Esta unidad se encuentra en contacto por falla e intrusión con el Pórfido Riolítico y andesitas de la formación Augusta Victoria.

2.2.1.3. Unidad Pórfido Riolítico

Se define como un cuerpo intrusivo hipabisal con una distribución más restringida, de la cual sus afloramientos se presentan en la parte más alta de la pared Norte del rajo Escondida, donde esta unidad intruye al CIFE y la formación Augusta Victoria.

Presenta fenocristales de biotita subhedrales de 2 mm de espesor, los que se encuentran alterados a clorita o sericita.

2.2.1.4. Brechas magmáticas-hidrotermales

Las brechas constituyen la unidad de roca que concentra la mayor cantidad de mineralización, tanto de mena como de ganga. En el yacimiento las brechas afectan a todas las unidades salvo las gravas. Equivalen aproximadamente al 5% del total de las rocas del yacimiento y son clasificadas en dos tipos: origen hidrotermal o ígneo. Ambas brechas pueden tener los mismos tipos de mineralización y su abundancia depende del proceso de hidrotermalismo involucrado [22].

2.2.1.5. Gravav

Corresponde a una secuencia de sedimentos continentales (areniscas y brechas sedimentarias) mal a moderadamente consolidadas, compuesta de fragmentos polimícticos, mal seleccionados, angulosos a subangulosos, de tamaños variables milimétricos a decimétricos cuya

composición corresponde a los tipos litológicos más antiguos. Presenta una matriz de composición variable, compuesta por detritos tamaño arena y cemento de yeso, carbonatos y sales indeterminadas.

Su disposición espacial es subhorizontal y en algunos sectores se observan niveles de mineralización exótica de cobre.

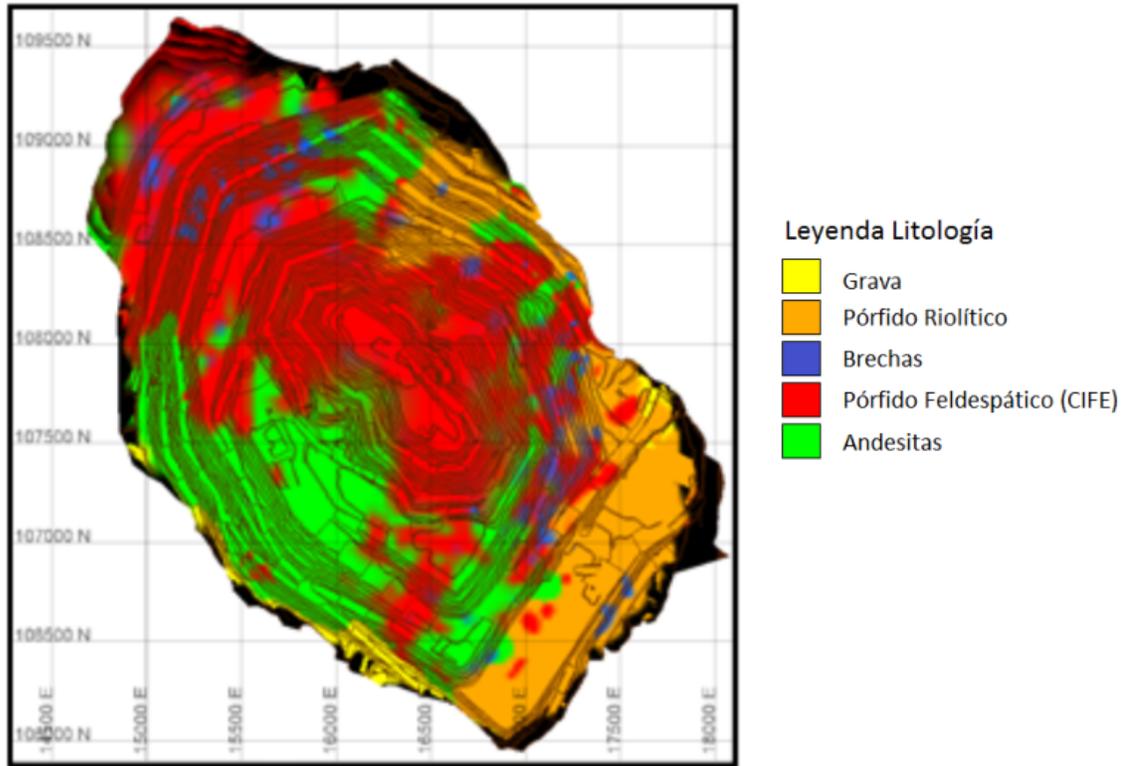


Figura 2.4: Litologías presentes en mina Escondida. Modificado de [1]

2.2.2. Alteración

El depósito Escondida presenta una clara zonación lateral y vertical de las alteraciones hidrotermales que afectan a las rocas (pág. 4). Se observa en la figura 2.5 que las alteraciones presentan cierta simetría en torno a un núcleo potásico, lo que concuerda con un pórfido cuprífero idealizado [19]. En este modelo se denota un centro potásico, en torno al cual se disponen las alteraciones filica, propilítica y argílica.

2.2.2.1. Alteración Potásica (KK)

La alteración potásica se encuentra presente como un núcleo, cubriendo una superficie relativamente menor y presente principalmente en la pared Oeste del rajo Escondida (figura 2.5). Afecta a rocas del CIFE y a las andesitas de la formación Augusta Victoria. Está caracterizada principalmente por la asociación feldespato potásico-cuarzo-biotita±sericita, en

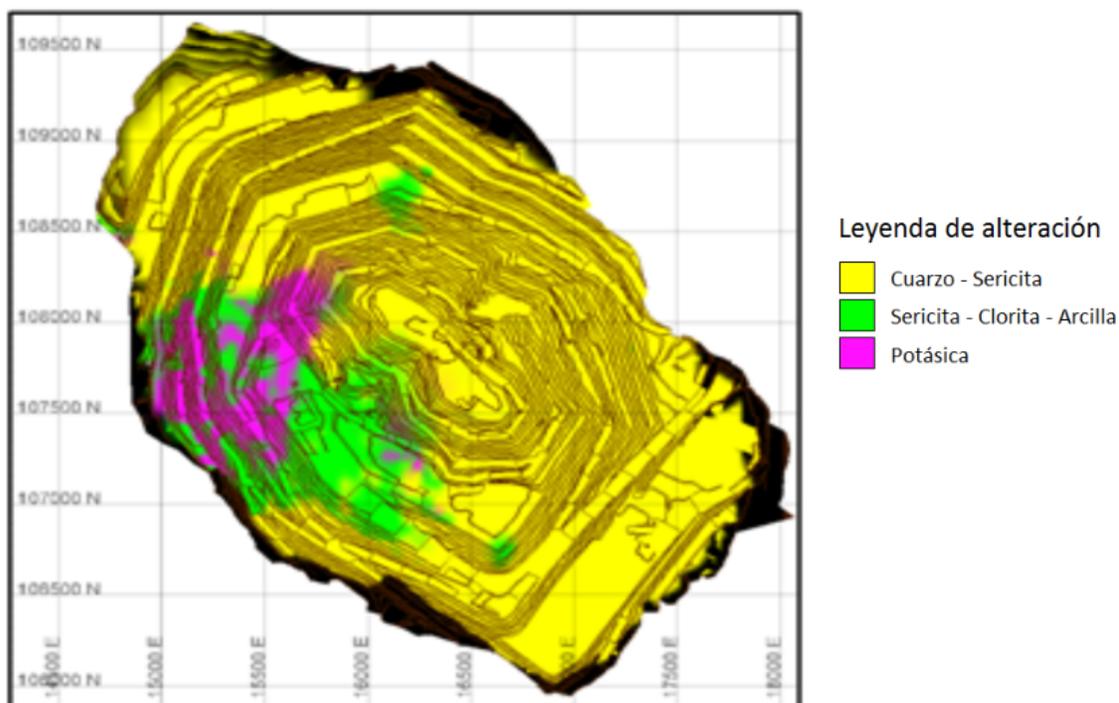


Figura 2.5: Alteraciones presentes en mina Escondida. Modificado de [1]

las que se pueden encontrar vetillas de feldespato potásico, las que en ocasiones se encuentran asociadas a cuarzo. Existe una disseminación de estos dos minerales más biotita en la matriz de rocas.

Por otro lado, en especial en las andesitas, se encuentra la alteración biotítica, la cual se caracteriza por una casi total alteración de los máficos a biotita. La asociación para esta etapa está representada por biotita-sericita y cuarzo y se interpreta como una etapa tardía de alteración potásica.

2.2.2.2. Alteración Sericita-Clorita-Arcilla (SCC)

La alteración Sericita-Clorita-Arcilla corresponde a la descripción clásica de alteración propilítica. Las biotitas y plagioclasas son reemplazadas por clorita-sericita-pirita y sericita-arcilla, respectivamente. Las arcillas asociadas a este tipo de alteración corresponden a caolinita y esmectita.

En Escondida afecta principalmente a las andesitas y en forma secundaria al CIFE.

2.2.2.3. Alteración Cuarzo-Sericita (QS,QSA)

Esta alteración presenta una relación directa con las zonas de más alto enriquecimiento mineral. En Escondida se encuentra principalmente en el CIFE, desde el techo de sulfuros hasta la base del enriquecimiento y es la que presenta mayor extensión en el rajo (figura 2.5).

2.2.2.4. Alteración Argílica

La alteración argílica se encuentra caracterizada en parte por la asociación cuarzo-caolinita, a la que se le denomina alteración argílica intermedia típica de esta etapa y es una alteración muy penetrativa, que en ocasiones oblitera en gran medida la roca original. Se distribuye principalmente en las paredes Norte y Este del rajo Escondida y se presenta junto a la alteración cuarzo-sericita y argílica avanzada.

2.2.2.5. Alteración Argílica Avanzada

La alteración argílica avanzada corresponde a una etapa hidrotermal tardía y se reconoce por exhibir la formación de cúmulos irregulares (motas) de color blanco intenso, las cuales presentan una asociación sericita, alunita, pirofilita y diásporo, insertas en una masa subtranslúcida de color gris, compuesta principalmente de sílice y sericita. En Escondida, esta alteración se distribuye en las partes altas de los sectores noroeste y sureste del yacimiento (figura 2.5).

2.2.3. Mineralización

La mineralización está directamente relacionada con la intrusión del CIFE [22]. Este complejo fue el que aportó la mineralización primaria y los fluidos hidrotermales responsables de la etapa de alteración hidrotermal, a la que se asocia el enriquecimiento secundario. La mineralización presenta un control estructural, principalmente alojado en fracturas, tanto en el caso de la mineralización hipógena como supérgena.

2.2.3.1. Mineralización hipógena

La mineralización hipógena en Escondida se puede clasificar en tres principales asociaciones minerales relacionadas a tres etapas hidrotermales [15]:

1. En la etapa de alteración potásica está la asociación magnetita-bornita-calcopirita con un contenido de sulfuros menor al 0.5 % en volumen.
2. En la zona de alteración sericita-clorita es característica la ocurrencia de vetillas de molibdenita y piritita intercrecidas con sericita, con un aumento de la cantidad de sulfuros a un 2 % en volumen.
3. La etapa hidrotermal tardía, que se caracteriza por la presencia de vetas polimetálicas con sulfuros como calcopirita, bornita, piritita lamelar, covelina, calcosina, enargita, esfalerita, galena y tenantita.

2.2.3.2. Mineralización supérgena

Esta mineralización puede ser representada en zona de lixiviados, zona de óxidos y zona de sulfuros secundarios (figura 2.6). La zona de lixiviados se encuentra definida por minerales de hierro oxidados tales como goethita, magnetita y jarosita, las cuales se distribuyen en la zona superior del rajo.

En la zona oxidada de cobre coexisten óxidos y sulfatos de cobre, entre ellos se destacan la brochantita y antlerita, y en menor grado atacamita, crisocola, pseudomalaquita, libetenita, turquesa, cuprita, cobre nativo, copper wad y copper pitch.

Bajo la zona oxidada de cobre se encuentra la zona de mixtos, la cual contiene mezclas de sulfuros de cobre, primarios y secundarios, con óxidos de cobre. En Escondida están principalmente representados por sulfatos de cobre hidratados (brochantita y antlerita), equivalentes a un 90 % del total de óxidos con menor proporción de atacamita, crisocola, malaquita, azurita, cuprita, cobre nativo y turquesa.

La zona de sulfuros secundarios corresponde a la ocurrencia dominante y/o continua de mineralización sulfurada de cobre de carácter secundario, tales como calcosina y covelina, localizada en el techo dominante de sulfuros y la zona primaria. Al interior de esta zona secundaria se distingue una clara zonación mineralógica vertical [22], con intensidad variable de enriquecimiento, representados desde los niveles superiores a inferiores por:

1. Calcosina + Pirita
2. Calcosina + Pirita (- Calcopirita)
3. Calcosina + Pirita (- Calcopirita - Covelina)
4. Pirita + Calcopirita + Covelina (- Calcosina)
5. Pirita + Calcopirita + Covelina (- Calcosina - Bornita)

Según esta zonación mineralógica vertical, los sulfuros secundarios se subdividen en dos: una subzona superior, denominada Alto Enriquecimiento (HE) (figura 2.6), que presenta como mena principal calcosina en pátina sobre pirita, en vetillas y en reemplazo y en forma subordinada y esporádica covelina y calcopirita. Una subzona inferior, denominada Bajo Enriquecimiento (LE), de la cual se distingue por la presencia continua de calcopirita, junto a covelina, pirita y menor cantidad de calcosina.

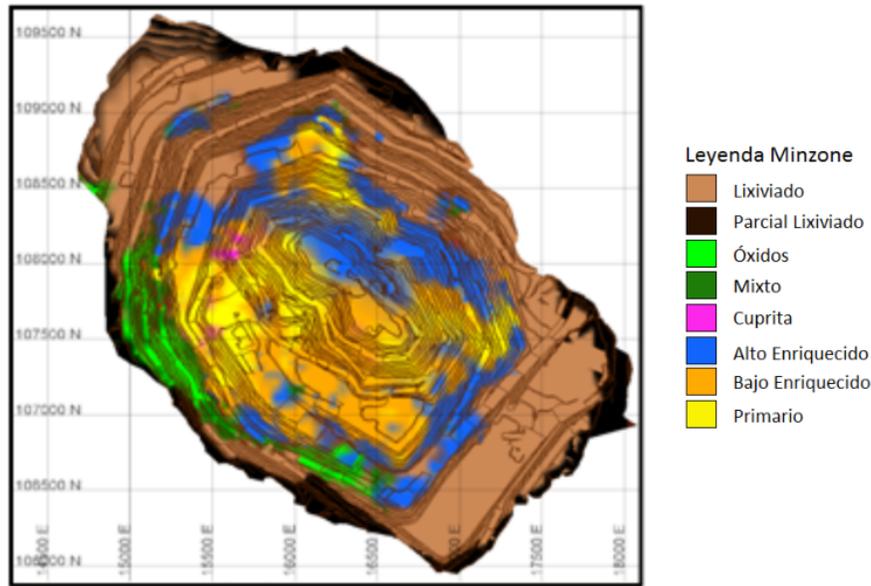


Figura 2.6: Mineralizaciones presentes en mina Escondida. Modificado de [1]

2.3. Flotación de minerales de cobre

Los datos de recuperación de cobre fueron obtenidos a partir de la etapa de flotación rougher dentro de un circuito de flotación de minerales de cobre, por lo que es necesario definir este concepto.

El proceso de flotación de minerales de cobre consiste en la concentración de minerales sulfurados de cobre, tales como la calcopirita y la bornita, en donde el material chancado y molido desde la operación minera, ya convertido en pulpa, se le añaden reactivos y es introducido en receptáculos llamados celdas de flotación, en las cuales se burbujea aire y se agita para mejorar el proceso, generando una pulpa con una alta concentración de cobre (aproximadamente 30 %).

Un circuito convencional de flotación se detalla en la Figura 2.7, donde se observan diversas etapas:

- Flotación primaria o *rougher*: Es la primera etapa de separación del mineral, donde el objetivo es maximizar la recuperación, generando un relave con una «baja» ley.
- Flotación de limpieza o *cleaner*: Destinada a incrementar la ley del concentrado. En esta etapa se genera el producto final de la planta.
- Flotación de barrido, o *scavenger*: Destinada a maximizar recuperación, sin embargo, trata relaves para evitar pérdidas.

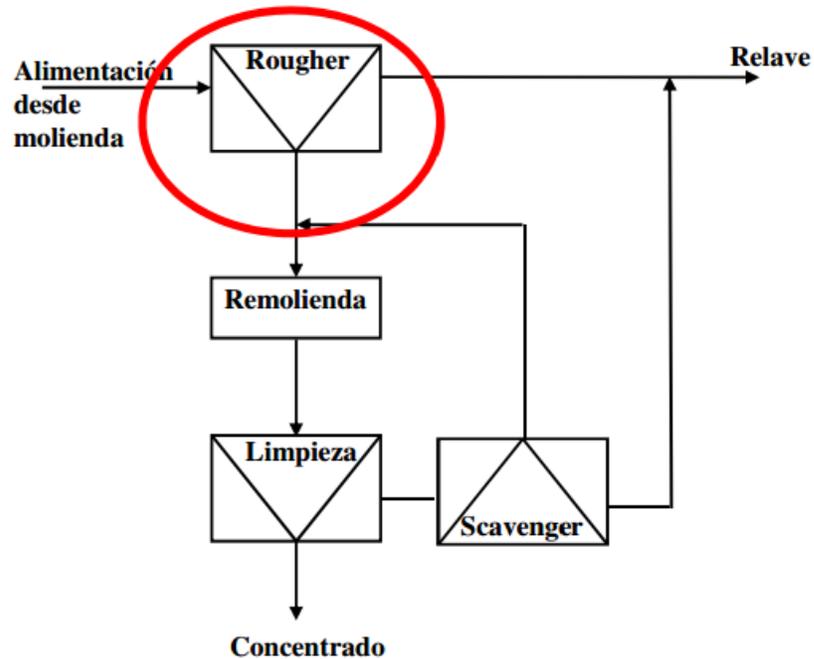


Figura 2.7: Circuito convencional de flotación de cobre, en donde se destaca la etapa rougher, fuente de los datos experimentales de recuperación. Fuente: [12]

Una de las variables que afecta la recuperación de cobre en el proceso de flotación, que será de relevancia para este trabajo, es el contenido de arcillas. Con la presencia de arcillas, tales como la *kaolinita* e *illita*, se provoca una pérdida en la recuperación de cobre debido a que se forma una capa de limo en la superficie del mineral y/o en las burbujas, que conlleva a un aumento del consumo de reactivo y traspaso de limo o fango al resto de las etapas de flotación [3].

2.4. Análisis multivariable

Los datos a estudiar en este trabajo de memoria presentan una gran cantidad de variables, de las cuales algunas pueden aportar información y otras no. El análisis multivariable apunta a interpretar las múltiples relaciones entre las variables presentes, así como buscar la simplicidad en su análisis. Entre las múltiples herramientas de análisis multivariable existentes actualmente, se dará énfasis en este trabajo a las técnicas de análisis de componentes principales y agrupamiento *k-Means*.

2.4.1. Estudio exploratorio de datos

En esta sección se describirán las herramientas básicas de estudio exploratorio de datos.

2.4.1.1. Histograma

Un histograma es una representación gráfica de una variable en forma de barras, donde se puede observar la frecuencia (o cantidad de datos) de la variable en clases, en general definidas como intervalos del mismo tamaño referentes al valor de la variable. Los histogramas son útiles para la detección de valores atípicos (outliers), además de identificar comportamiento bimodal o multimodal en las variables y permitir visualizar la distribución de los valores (rango, simetría, forma, etc.).

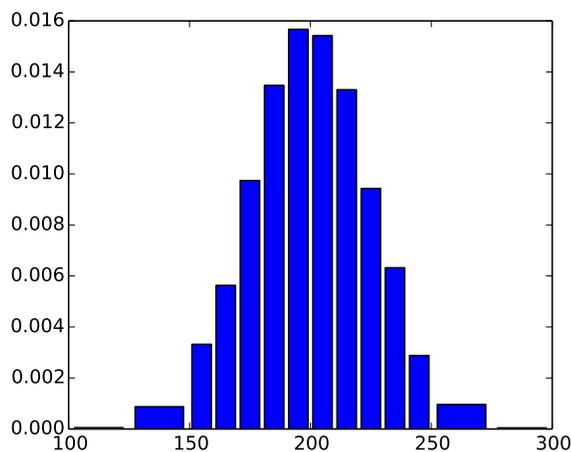


Figura 2.8: Ejemplo de histograma de frecuencias

2.4.1.2. Diagramas de caja (*boxplot*)

Un diagrama de caja o *boxplot* es un gráfico que permite resumir algunas características de la distribución de una variable, tales como su simetría y dispersión, en un gráfico de una dimensión. Los elementos descritos en un *boxplot* son los cuartiles superior e inferior, la media, el mínimo y máximo de los datos.

Un *boxplot* (figura 2.9) también permite visualizar outliers o datos atípicos en función del rango intercuartil de los datos (RIC), el cual se define como la diferencia entre los cuartiles superior e inferior ($RIC = F_U - F_L$). Un dato se define como atípico si no pertenece al intervalo $[F_L - 1.5 \times RIC, F_U + 1.5 \times RIC]$.

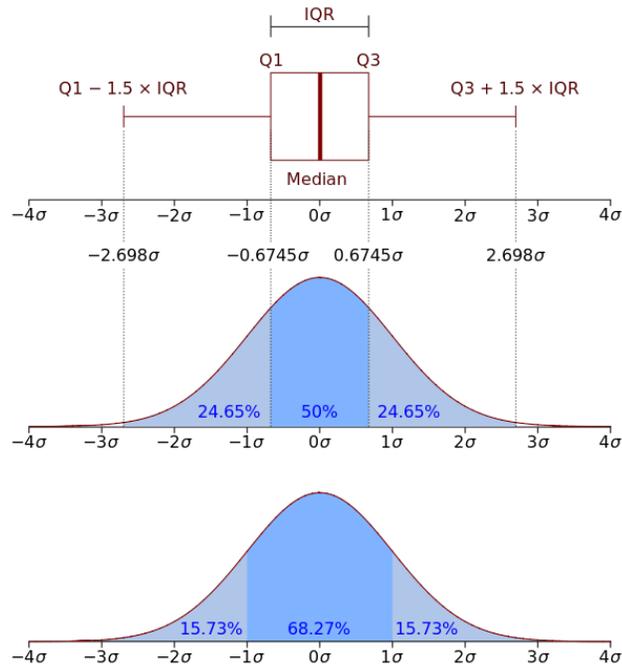


Figura 2.9: Ejemplo de un diagrama de caja (boxplot)

2.4.2. Coeficiente de correlación lineal de Pearson

El coeficiente de correlación lineal de Pearson es un índice que mide la relación lineal entre dos variables cuantitativas. Se simboliza con la letra $\rho_{x,y}$ y se expresa como:

$$\rho_{x,y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Donde:

- σ_{XY} es la covarianza de (X, Y)
- σ_X es la desviación típica de la variable X
- σ_Y es la desviación típica de la variable Y

El valor del índice de correlación de Pearson varía en el intervalo $[-1, 1]$, interpretable de la siguiente manera:

- Si $\rho_{x,y} = 1$, existe una correlación positiva perfecta, o en otras palabras, existe una relación lineal entre las dos variables.
- Si $0 < \rho_{x,y} < 1$, existe una correlación positiva.
- Si $\rho_{x,y} = 0$, no existe relación lineal. Sin embargo no implica que no exista dependencia entre las variables ya que ésta puede ser no lineal.
- Si $-1 < \rho_{x,y} < 0$, existe una correlación negativa.

- Si $\rho_{x,y} = -1$, existe una correlación negativa perfecta, lo que indica una relación lineal inversa entre las dos variables.

Un ejemplo gráfico del índice de correlación de Pearson se puede apreciar en la figura 2.10.

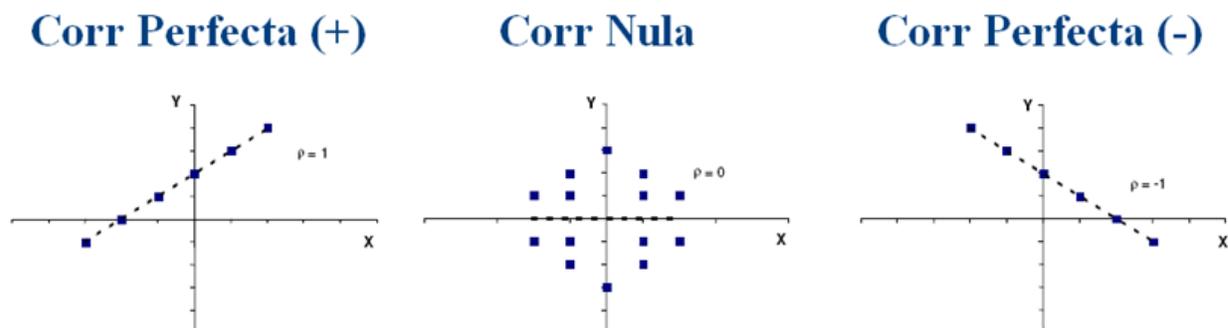


Figura 2.10: Ejemplo del índice de correlación de Pearson. Fuente: [8]

2.4.3. Matriz de correlación

La matriz de correlación es una matriz cuadrada simétrica, constituida por los coeficientes de correlación de Pearson de cada pareja de variables, en la cual se tendrán unos en la diagonal de la matriz y el valor de correlación $\rho_{x,y}$ en los elementos no diagonales. Un ejemplo se puede observar en la tabla 2.1.

Tabla 2.1: Ejemplo de matriz de correlación

Variables	Cd	Co	Cr	Cu	Ni	Pb	Zn
Cd	1.00	0.26	0.58	0.15	0.49	0.22	0.62
Co	0.26	1.00	0.48	0.19	0.74	0.16	0.44
Cr	0.58	0.48	1.00	0.21	0.71	0.26	0.61
Cu	0.15	0.19	0.21	1.00	0.22	0.82	0.66
Ni	0.49	0.74	0.71	0.22	1.00	0.27	0.59
Pb	0.22	0.16	0.26	0.82	0.27	1.00	0.67
Zn	0.62	0.44	0.61	0.66	0.59	0.67	1.00

2.4.4. Análisis de componentes principales

Uno de los principales problemas en el análisis de datos multivariantes es la reducción de la dimensionalidad, en otras palabras, si es posible describir con precisión las p variables con un subconjunto de variables $k < p$ de ellas, entonces se habrá reducido la dimensionalidad del problema con una pequeña pérdida de información.

El análisis de componentes principales (PCA) es una técnica estadística multivariable, que permite transformar un conjunto de variables originalmente correlacionadas entre sí,

en un conjunto sintético de variables no correlacionadas llamadas factores o «componentes principales» [9], las cuales no contienen información repetida. Las nuevas variables son combinaciones linealmente independientes de las variables originales, ordenadas de acuerdo a la representación de dispersión respecto a la nube total de información recogida de las muestras.

2.4.4.1. Algoritmo

Sea un conjunto de variables (x_1, \dots, x_p) , cada uno con el mismo número de muestras n , representado de la siguiente forma:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

La matriz X constituye la tabla de datos u observaciones, la cual representa el total de la información. A partir de esta matriz se calcula un nuevo conjunto (C_1, \dots, C_p) no correlacionados entre sí, en la cual sus varianzas sean decrecientes. Cada elemento $C_j (j = 1, \dots, p)$ constituye una combinación lineal de las variables originales (x_1, \dots, x_p) :

$$C_j = a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jp}x_p$$

$$C_j = a'_j x$$

donde

$$a'_j = (a_{1j}, a_{2j}, \dots, a_{pj})$$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$$

a'_j es un vector de constantes y representa el peso de las variables en cada componente. Estas explican una parte de la varianza total, apuntando a encontrar k componentes que expliquen casi toda la varianza en la nube de información.

Condiciones para obtener la primera componente principal:

1. Para mantener la ortonormalidad de la transformación se debe imponer que:

$$a_j' a_j = \sum_{k=1}^p a_{kj}^2 = 1$$

2. La varianza en la primera componente principal C_1 debe ser la máxima en la nube de datos.

El primer componente principal representa la mayor varianza observada, el segundo componente principal se obtiene calculado los valores de a_2 de tal manera que C_1 y C_2 estén no correlacionados (ortogonales). La obtención de los siguientes componentes principales se obtienen de la misma manera, cada uno con varianza menor que el anterior.

2.4.4.2. Selección del número de componentes

La cantidad total de componentes principales es igual al número de variables presentes en los datos. Sin embargo, al elegir más de 3 componentes se complica la interpretación gráfica y la aplicación de algoritmos de clustering sobre ellos, por lo que es relevante determinar cuántas componentes elegir con la menor pérdida de varianza explicada posible. Para determinar esta cantidad de componentes se tienen dos criterios:

- Seleccionar componentes de tal forma que la proporción de la varianza acumulada sea igual o mayor a un criterio definido a priori, como por ejemplo 80 % o 90 %. Sin embargo, este criterio no debe ser aplicado a ciegas, ya que perfectamente la primera componente puede cumplir este criterio, y se termina descartando otras componentes que pueden representar la verdadera forma de las variables.
- A través de un *gráfico de sedimentación*, el cual representa los factores en el eje de las abscisas y los valores propios en las ordenadas. Para elegir la cantidad apropiada de componentes principales se debe distinguir un punto de inflexión, o «codo», que indica a partir de cual componente se obtienen bajas varianzas, por lo tanto se deben seleccionar las componentes hasta el punto de inflexión o incluyendo la siguiente, dependiendo del criterio del investigador. En el ejemplo de la Figura 2.11, el punto de inflexión se ubica en la segunda componente, sin embargo, dependiendo del criterio del analista también puede seleccionar hasta la tercera componente.

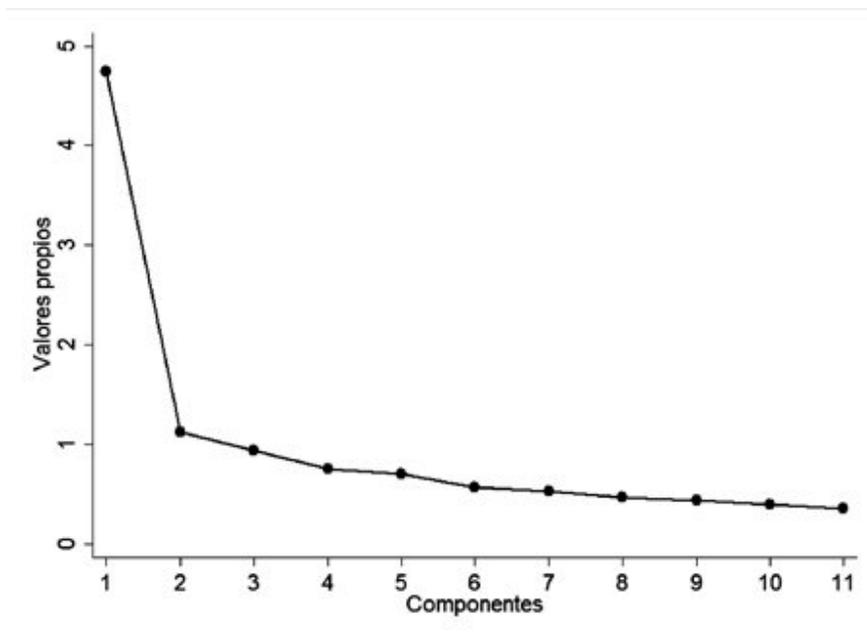


Figura 2.11: Ejemplo de gráfico de sedimentación.

2.4.5. Clustering

El agrupamiento o *clustering* se define como una metodología que permite dividir n muestras en k grupos. El objetivo del agrupamiento es maximizar la similitud entre las muestras de un mismo grupo y además minimizar la similitud entre las muestras de grupos distintos.

En este trabajo se utilizaron y compararon dos algoritmos de clustering, uno del tipo no jerárquico o de partición (*k-Means*) y uno jerárquico (*Ward*). El detalle de cada uno de estos algoritmos se mostrará a continuación.

2.4.5.1. k-Means

El primer algoritmo a detallar es *k-Means*, el cual es un método del tipo no jerárquico o de partición. Este algoritmo es extremadamente eficiente y es comúnmente utilizado para clustering de una gran cantidad de datos. Sin embargo, requiere especificar el número de clusters a definir, por lo que la selección de esta cantidad dependerá del criterio del evaluador.

El algoritmo *k-Means* se inicializa tanto asignando datos a uno de los K clusters predefinidos y después calculando sus K centroides. Estos centroides pueden ser obtenidos aleatoriamente o asignados arbitrariamente. A continuación busca minimizar el valor de ESS (Explained sum of squares, ecuación (2.1)) reasignando datos en clusters específicos. Este procedimiento se realiza hasta encontrar un valor mínimo de ESS [13].

Algoritmo

1. Input: $T = \{x_i, i = 1, \dots, n\}$, con K número de clusters.
2. Pre-especificar los K centroides aleatoriamente. $\bar{x}_k, k = 1, \dots, K$
3. Calcular el cuadrado de la distancia Euclídeana de cada ítem con respecto al centroide de su clúster.

$$ESS = \sum_{k=1}^K \sum_{c(i)=k} (x_i - \bar{x}_k)^T (x_i - \bar{x}_k) \quad (2.1)$$

donde \bar{x}_k es el centroide k -ésimo y $c(i)$ es el cluster que contiene x_i

4. Reasignar cada ítem al centroide de clúster más cercano de forma que ESS disminuya. Actualizar los centroides después de cada re-asignación.
5. Repetir pasos 3 y 4 hasta que no exista re-asignación de los datos.

La solución obtenida es un mínimo local de ESS, no global, además se recomienda que el posicionamiento de los centroides sea aleatorio [10]. Un ejemplo se puede apreciar en la figura 2.12, donde se cuenta con dos variables, las cuales son agrupadas en 5 conjuntos optimizando desde A a D, redistribuyendo los centroides.

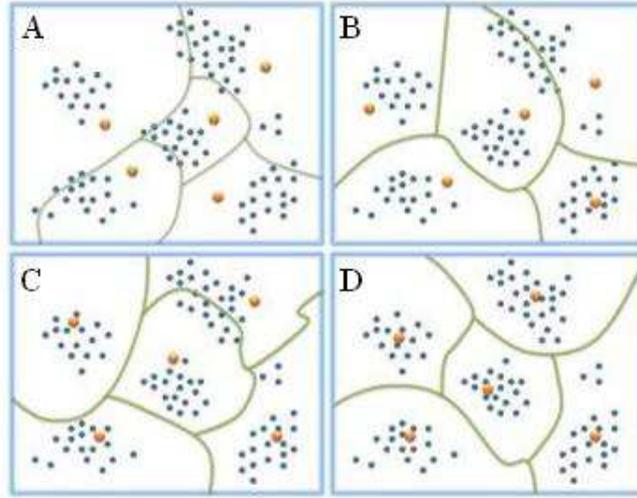


Figura 2.12: Ejemplo de agrupamiento k -Means. Fuente: [11]

2.4.5.2. Ward

El método de Ward [23] es un criterio aplicado al análisis de cluster jerárquico aglomerativo (Anexo B.1). Este algoritmo presenta la particularidad de que no utiliza la distancia entre clusters para agrupar objetivos. En cambio, utiliza la suma de cuadrados dentro del cluster (Ecuación 2.2) para determinar qué par de grupos serán unidos en cada paso del algoritmo.

$$SSE = \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (2.2)$$

donde y_{ij} es el j -ésimo objeto en el cluster i -ésimo y n_i es el número de objetos en el i -ésimo cluster.

2.4.5.3. Coeficiente de silueta

Es importante evaluar el desempeño de estos algoritmos de clustering. Sin embargo ésta no es una tarea trivial ya que no se tienen los resultados reales o esperados para poder realizar comparaciones. Un método de evaluación es una métrica llamada *coeficiente de silueta* [18], que cuantifica cuan bien están definidos los clusters. Este coeficiente se define para cada valor y presenta dos parámetros:

- a : La distancia media entre un punto y todos los otros puntos en la misma clase.
- b : La distancia media entre un punto y todos los otros puntos en el siguiente cluster más cercano.

El coeficiente de silueta para un punto se define como:

$$s = \frac{b - a}{\max(a, b)}$$

El coeficiente de silueta para un conjunto de datos se obtiene con la media de todos los coeficientes para cada punto. Un coeficiente con un valor de -1 indica un clustering incorrecto y +1 un clustering con perfecta clasificación. Valores nulos indican sobre posicionamiento de clusters y valores altos indican alta densidad en los clusters, indicando buena separación entre ellos.

2.4.5.4. Matriz de confusión

Una matriz de confusión tiene como objetivo evaluar la calidad de clasificación de algún método con respecto a un resultado esperado o real [17]. Los elementos en la diagonal representan el número de puntos para los cuales la clasificación predicha es igual a la verdadera, en cambio, los elementos fuera de la diagonal son los que están incorrectamente etiquetados por el clasificador. A medida que los valores en la diagonal sean mayores, más exacta es la predicción.

Un ejemplo de una matriz de confusión se detalla en la Tabla 2.2.

Tabla 2.2: Ejemplo de matriz de confusión

		Predicho	
		ABRIR	CERRAR
Esperado o real	ABRIR	TP	FP
	CERRAR	FN	TN

donde:

- *TP*: son los casos que pertenecen a esa clase y el clasificador los definió en ella.
- *FM*: son los casos que pertenecen a la clase y el clasificador los definió incorrectamente en otra.
- *FP*: son los casos que no pertenecen a la clase, pero el clasificador los asoció a ella.
- *TN*: son los casos que no pertenecen a la clase y el clasificador determinó lo mismo.

2.5. Modelo predictivo

El análisis de regresión es usado para explicar o modelar la relación entre una variable Y , llamada la de *respuesta, salida o dependiente*, y una o varias variables *predictivas, independientes o explicativas*, X_1, \dots, X_p . Cuando $p = 1$ la regresión es del tipo simple y cuando $p > 1$ se le llama regresión multivariable. Esta última será la utilizada para obtener un modelo predictivo de la recuperación de cobre.

El objetivo del análisis de regresión es básicamente:

1. Predicción de futuras observaciones.
2. Evaluación del efecto, o la relación, de las variables explicativas sobre la de respuesta.
3. Proveer una descripción general de la estructura de datos.

En el contexto de este trabajo se realizará una regresión multilineal sobre los datos, que permita predecir la recuperación de cobre por flotación rougher en función de las variables presentes. Sin embargo existen una gran cantidad de variables predictivas, por lo que el modelo tenderá al sobreajuste. Para evitar este problema, se realizará el análisis de regresión con las variables relevantes, las cuales serán seleccionadas bajo la metodología propuesta en la siguiente sección.

2.5.1. Selección de variables

El objetivo de la selección de variables es elegir el «mejor» conjunto de variables predictivas. Este conjunto debe ser elegido bajo las siguientes premisas:

1. El conjunto de variables debe representar a los datos de la manera más simple posible, en la que las variables predictivas redundantes deben ser eliminadas. El modelo más pequeño (menor cantidad de variables) que se ajuste de manera suficiente a los datos es mejor.
2. Variables predictivas innecesarias solo agregarán ruido a la estimación a realizar.
3. Se evita la colinealidad, causada al usar muchas variables al realizar la regresión.

La metodología aplicada para la selección de variables será a través de su coeficiente de correlación de Pearson (Sección 2.4.2) con respecto a la recuperación de cobre. En primer lugar

se omitirán las variables irrelevantes, en otras palabras, las que presenten menor coeficiente de correlación tanto positivo como negativo. La decisión sobre qué valores de este coeficiente se usarán dependerá del criterio y conocimiento del investigador en base a los resultados obtenidos. Una vez removidas las variables irrelevantes, se deben eliminar las que presenten redundancia entre sí. Una variable se define como redundante cuando presenta una correlación igual o mayor a 0.94 con respecto a otra, eliminando la que sea menos abundante (menor media). Este valor fue obtenido a partir del trabajo de Boisvert et al (2013) [2], determinado en base a un análisis de sensibilidad no detallado por los autores.

2.5.2. Regresión multilínea

Existen diversos algoritmos de regresión multilínea, que además de calcular los coeficientes asociados a las variables explicativas y el intercepto, también seleccionan un subconjunto de variables óptimas a usar para el modelo de regresión. El algoritmo que se utilizará en este trabajo es LARS Lasso, el cual es rápido computacionalmente y permite realizar análisis de sensibilidad al imponer una restricción en la suma de los valores absolutos de los coeficientes de regresión [7].

El algoritmo se define de la siguiente forma:

2.5.2.1. Algoritmo

LARS Lasso ajusta un modelo lineal del tipo:

$$\hat{y} = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (2.3)$$

donde y es la variable de respuesta que se quiere predecir, β_0 es una constante llamada *intercepto*. x_i son las n variables explicativas y β_i los n coeficientes que ponderan las variables explicativas.

Bajo el criterio:

$$\min \sum_i (y_i - \hat{y})^2 \quad (2.4)$$

tal que:

$$\sum_j^n |\beta_j| \leq s \quad (2.5)$$

La primera suma es sobre la cantidad total de datos. El valor s es considerado un parámetro de ajuste, el cual al ser suficientemente grande, el algoritmo LARS Lasso queda reducido a un simple modelo predictivo hecho a través de mínimos cuadrados. Sin embargo, al elegir un valor adecuado de s se pueden generar modelos multilineales reducidos donde algunos coeficientes β_j pueden ser cero, en otras palabras, la elección de s determinará cuantas variables quedarán en el modelo multilineal.

Un ejemplo se puede observar en la Figura 2.13, donde a medida que aumenta el valor de s (eje de la abscisa), la cantidad de variables y sus coeficientes aumentan, por lo que es importante realizar un análisis de sensibilidad con respecto a este valor.

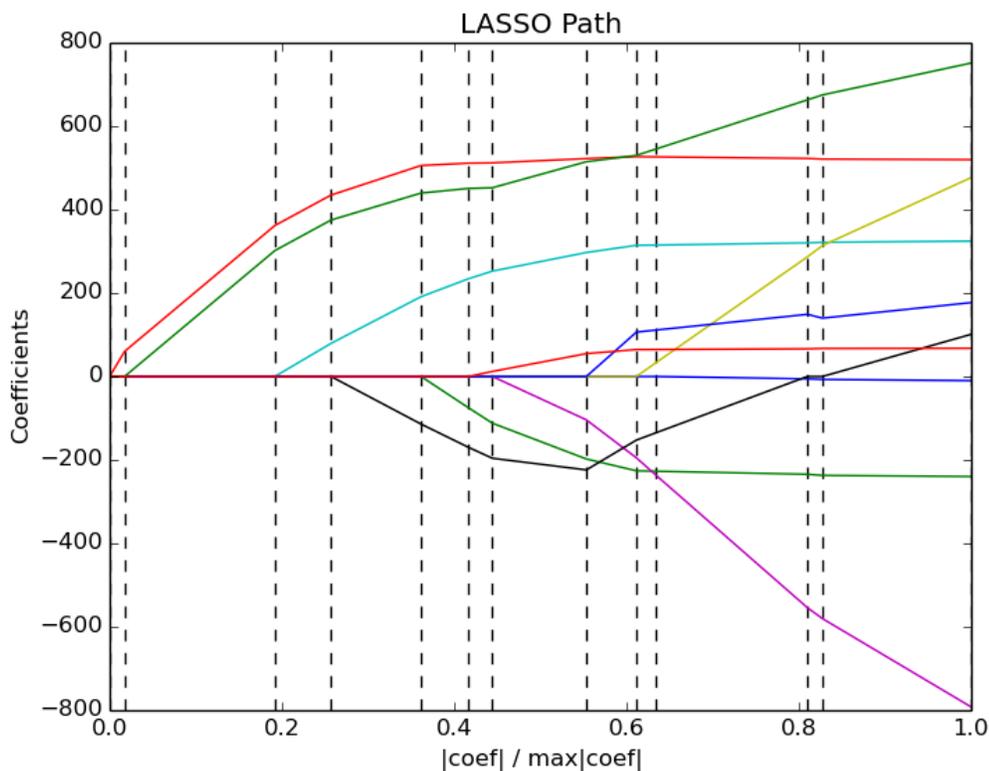


Figura 2.13: Ejemplo de algoritmo LARS Lasso. En el eje de las abscisas está el coeficiente s y en eje de las ordenadas el coeficiente de la variable predictiva.

Es sumamente importante destacar que en el paquete de software *scikit-learn*, el valor el coeficiente s ingresado como entrada presenta un significado diferente al detallado acá. En otras palabras, un valor bajo ($s = 0$) incluye todas las variables dentro del modelo predictivo multilineal.

2.5.3. Coeficiente de determinación

El coeficiente de determinación (R^2), o equivalente al coeficiente de correlación de Pearson (Sección 2.4.2) al cuadrado, es una medida adimensional que permite evaluar la calidad de ajuste del modelo multilineal a los datos. Se define de la siguiente forma:

$$R^2 = \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2}$$

donde:

- \bar{Y} : Media muestral de los datos.
- \hat{Y}_i : Dato i real.
- Y_i : Dato i esperado.

Un valor de R^2 igual a 1 indica un ajuste lineal perfecto, en otras palabras, que la variable es explicada totalmente por el modelo de regresión. El valor cero indica la no representatividad del modelo lineal.

Capítulo 3

Metodología

A continuación se presentará la metodología propuesta para determinar dominios geometalúrgicos, junto con la generación de un modelo predictivo de recuperación de cobre por flotación rougher.

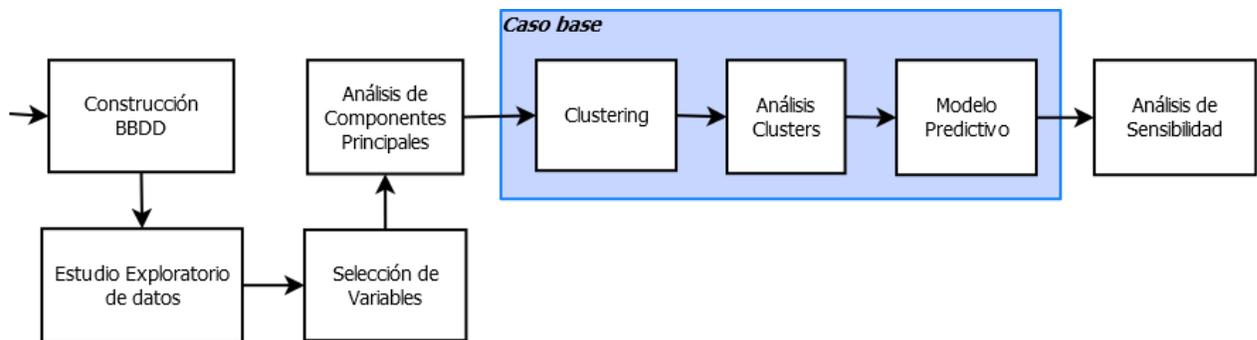


Figura 3.1: Esquema de la metodología a implementar

3.1. Construcción base de datos

En primer lugar se mencionarán las bases de datos disponibles, las que serán unidas en una base de datos única que es con la que se realizará el estudio. Posteriormente se adicionarán variables sintéticas, las cuales indican relaciones entre variables, identificando asociaciones minerales relacionadas con alteraciones.

3.2. Estudio exploratorio de datos

A partir de la base de datos construida se realizará un estudio exploratorio, de la cual se recopilarán estadísticas básicas tanto generales como para cada variable geológica (litologías,

alteraciones y mineralizaciones). Las variables a analizar serán las leyes de cada elemento presente y la recuperación de cobre.

3.3. Selección de variables

Previo a la definición de los dominios geometalúrgicos a través de análisis de componentes principales (PCA) y clustering, es necesario reducir la cantidad de variables a usar en el estudio, ya que al realizar PCA sobre una gran cantidad de variables generará componentes principales que explicarán un bajo porcentaje de la varianza total.

La metodología a usar para la selección de variables relevantes será a través del coeficiente de correlación de Pearson. Para ello se calculará este coeficiente entre las leyes de elementos y la recuperación de cobre. Una vez realizado se eliminarán del estudio las variables de leyes que presenten una baja correlación con respecto a la recuperación, tanto positiva como negativa. El criterio para elegir estos valores dependerán de los conocimientos del investigador a cargo en base a los resultados obtenidos a partir de los datos. Una vez removidas las variables no relevantes, se deben también eliminar las redundantes: una variable se define como redundante cuando presenta una correlación igual o mayor a 0.94 con respecto a otra, eliminando la que sea menos abundante (menor media) [2].

Una vez eliminadas las variables no relevantes y las redundantes se procederá a hacer la definición de dominios a través de PCA y clustering.

3.4. Análisis de componentes principales

Una vez seleccionadas las variables a utilizar en el estudio, se realizará un análisis de componentes principales, previa normalización de las variables, con la finalidad de reducir aún más la dimensionalidad de los datos a un espacio de 2 ó 3 dimensiones dependiendo de las componentes seleccionadas y el porcentaje de la varianza total que explican en conjunto. La elección de la cantidad de componentes principales a utilizar se realizará en base a dos criterios: el primero utilizando el gráfico de sedimentación obtenido y en segundo lugar considerar un porcentaje total de varianza explicada mínimo, el cual será de un 90 %.

3.5. Caso base

Una vez realizada la selección de las variables relevantes y el análisis de componentes principales, se procederá a realizar un caso base, en donde se estudiarán los resultados obtenidos al calcular 4 clusters, número el cual fue elegido para ejemplificar la utilización de la metodología propuesta y no guarda relación con la cantidad de clusters óptimos. Estos dominios serán obtenidos utilizando las herramientas de agrupamiento disponibles, para luego ser analizados y comparados entre sí, para finalmente calcular un modelo predictivo a partir

de los datos contenidos en cada cluster definido y establecer conclusiones con respecto a los resultados.

3.5.1. Clustering

En esta etapa se realizará clustering o agrupamiento sobre las componentes principales seleccionadas, con la finalidad de identificar los dominios geometalúrgicos a estudiar. Los algoritmos de agrupamiento a utilizar serán k-Means y Ward, siendo estos métodos de agrupamiento por particionamiento y de tipo jerárquico respectivamente.

3.5.2. Análisis de clusters

Una vez obtenidos los clusters (dominios) a través de cada método, se procederá a comparar los resultados de cada uno con la finalidad de ver cual método agrupa mejor los datos y qué tan semejantes son los dominios obtenidos por un método con respecto al otro.

El criterio a usar para ver cual método agrupa mejor es el coeficiente de silueta, que a mayores valores indica grupos mejor definidos. En cuanto a comparar los dominios obtenidos bajo cada método, este se realizará a través de una matriz de confusión, que indica porcentajes de pertenencia de datos entre un cluster definido por un método, con respecto al mismo cluster generado por el otro método.

3.5.3. Modelo predictivo

Una vez definidos los clusters, se realizará un modelo predictivo de la recuperación de cobre en cada uno de los dominios obtenidos. El modelo predictivo será una regresión multilineal obtenida a través del algoritmo LARS Lasso, el cual permite seleccionar variables sobre el modelo predictivo al incorporar un parámetro de sensibilidad en la obtención de los coeficientes de regresión. Este parámetro permite realizar un *trade-off* entre la cantidad de variables a usar para la regresión y el coeficiente de determinación asociado (R^2), labor la cual dependerá del criterio del modelador a cargo. Finalmente, se detallarán y analizarán los resultados obtenidos al variar este parámetro.

3.6. Análisis de sensibilidad

Una vez hecho el estudio del caso base se realizará un análisis de sensibilidad, que tiene como objetivo identificar cuan bien son separadas las variables geológicas al variar la cantidad de dominios calculados. Esta actividad se realizará comparando proporciones de litologías, alteraciones y mineralizaciones entre cada dominio obtenido por los métodos de k-Means y Ward. Este procedimiento se hará para 4, 5 y 6 clusters por cada método.

Finalmente, en las conclusiones se comprobará que la elección de cuantos dominios (clusters) elegir dependerá del criterio del investigador, interpretando en base a los resultados cuan bien agrupan variables geológicas sin que existan dominios redundantes entre sí.

Capítulo 4

Caso de estudio

En este trabajo de título se usarán datos pertenecientes a Minera Escondida. Se trabajará esencialmente con tres bases de datos, la primera conteniendo información geoquímica, indicando el contenido de los elementos químicos presentes y su posición en el espacio. La segunda base de datos aporta la información geológica, incluyendo específicamente litología, alteración y mineralización. Finalmente, la tercera base de datos presenta información de recuperación rougher a partir de ensayos de laboratorio.

En base a estas tres bases de datos se construirá una base de datos única que incluya todas las variables relevantes: posición espacial, geoquímica (leyes), variables geológicas (litología, alteración y mineralización) y la recuperación de cobre. Las especificaciones de cada una de estas tres bases de datos y el cómo se construyó la base de datos final se detallará a continuación.

4.1. Bases de datos originales

4.1.1. Geoquímica

Esta base de datos contiene 15409 datos y presenta las siguientes variables:

- Identificación
 - *Bhid*: identificador único del sondaje.
- Posición
 - *From*: Posición «desde» del sondaje. En metros.
 - *To*: Posición «hasta» del sondaje. En metros.
 - *Este*: Posición en el eje este. En metros.
 - *Norte*: Posición en el eje norte. En metros.
 - *Cota*: Posición en la cota. En metros.
- Leyes (51 elementos)

Tabla 4.1: Elementos presentes en la base de datos de geoquímica

Elemento	Nombre	Elemento	Nombre	Elemento	Nombre
Ag	Plata	Ge	Germanio	S	Azufre
Al	Aluminio	Hf	Hafnio	Sb	Antimonio
As	Arsénico	Hg	Mercurio	Sc	Escandio
Au	Oro	In	Indio	Se	Selenio
B	Boro	K	Potasio	Sn	Estaño
Ba	Bario	La	Lantano	Sr	Estroncio
Be	Berilio	Li	Litio	Ta	Tantalio
Bi	Bismuto	Mg	Magnesio	Te	Telurio
Ca	Calcio	Mn	Manganeso	Th	Torio
Cd	Cadmio	Mo	Molibdeno	Ti	Titania
Ce	Cerio	Na	Sodio	Tl	Talio
Co	Cobalto	Nb	Niobio	U	Uranio
Cr	Cromo	Ni	Níquel	V	Vanadio
Cs	Cesio	P	Fósforo	W	Wolframio
Cu	Cobre	Pb	Plomo	Y	Itrio
Fe	Hierro	Rb	Rubidio	Zn	Zinc
Ga	Galio	Re	Renio	Zr	Circonio

Todas las unidades de los elementos en la Tabla 4.1 están en partes por millón (ppm), a excepción de *Al*, *Ca*, *Cu*, *Fe*, *K*, *Mg*, *Na* y *S* que están expresadas en porcentaje (%)

4.1.2. Geología

Esta base de datos contiene 35934 datos y presenta las siguientes variables:

- Identificación
 - *Bhid*: identificador único del sondaje.
- Posición
 - *From*: Posición «desde» del sondaje. En metros.
 - *To*: Posición «hasta» del sondaje. En metros.
- Geología
 - *Lito*: Litología
 - *Alte*: Alteración
 - *Minz*: Mineralización

4.1.3. Recuperación

Esta base de datos contiene 2457 datos y presenta las siguientes variables:

- Identificación

- *Bhid*: identificador único del sondaje.
- Posición
 - *From*: Posición «desde» del sondaje. En metros.
 - *To*: Posición «hasta» del sondaje. En metros.
 - *Length*: Largo del sondaje, corresponde a la diferencia entre To y From. En metros.
- Recuperación
 - *rec30*: Recuperación en flotación rougher bajo una granulometría # 30 Tyler¹, obtenida en laboratorio. En porcentaje.

4.2. Base de datos final

Para realizar un estudio adecuado de los datos se requiere relacionar las leyes de los elementos presentes con sus asociaciones geológicas y la recuperación de cobre. Para ello se creará una base de datos a partir de la unión de los datos de geoquímica, geología y recuperación mostrados anteriormente. La base de datos final es creada cruzando las variables *bhid*, *from* y *to* de cada base de datos, obteniendo una base con 1916 datos, la cual presenta las siguientes variables:

- Identificación
 - *Bhid*: identificador único del sondaje.
- Posición
 - *From*: Posición «desde» del sondaje. En metros.
 - *To*: Posición «hasta» del sondaje. En metros.
 - *Este*: Posición en el eje este. En metros.
 - *Norte*: Posición en el eje norte. En metros.
 - *Cota*: Posición en la cota. En metros.
- Geología
 - *Lito*: Litología
 - *Alte*: Alteración
 - *Minz*: Mineralización
- Recuperación
 - *rec30*: Recuperación en flotación rougher bajo una granulometría # 30 Tyler, obtenida en laboratorio. En porcentaje.
- Leyes: Los elementos presentes en la Tabla 4.1

Finalmente, esta base de datos contiene todas las variables relevantes y es la que será usada para realizar la definición de dominios geometalúrgicos junto con los modelos predictivos.

¹Aproximadamente 550 micrones.

4.2.1. Variables sintéticas

A la base de datos final se le adicionaron 13 variables sintéticas obtenidas a partir del trabajo de GeoAV Geochemistry & Exploration S.A. [21], las cuales son utilizadas a partir de la información contenida en la base de datos. Estas variables sintéticas son operaciones aritméticas entre variables geoquímicas que permiten identificar intensidades de diversas alteraciones, de las cuales, la multiplicación realza elementos cuando están en una alta concentración y los deprime cuando su concentración es baja. La división permite apreciar como un elemento varía con respecto a otro. La suma indica composiciones de elementos.

Las variables sintéticas añadidas son las siguientes:

Variables indicadoras

- Cu/S : Indica la presencia de sulfuros de cobre.
- K/Al : Razón entre potasio y aluminio. [2]

Alteración fílica

- $K * Al$: Realza la aparición de sericita/moscovita.
- $(K * Na)/Al$: En la estructura molecular de la moscovita quedan átomos de sodio atrapados en las intercapas, reemplazando al potasio por intercambio catiónico. La división por aluminio es agregada debido a que está presente en minerales de distintas alteraciones tales como la biotita y la alunita.
- Na/Al : Estudia el sodio con relación al aluminio.
- $(Al + K)/(Na + Ca + Mg)$: Cuantifica el proceso de intercambio entre los cationes aluminio y potasio con el grupo de cationes sodio, calcio y magnesio, los cuales presentan comportamiento semejante.
- $(Al + K + Na)/(Ca + Mg)$: Análogo al ítem anterior.

Alteración propilítica

- $(Ca + Na)/(K + Al)$: Destaca el proceso de intercambio entre los cationes de calcio y sodio con los cationes de potasio y aluminio.

Alteración argílica

- $Al/(Na + Ca + K)$: Cuantifica un enriquecimiento de aluminio asociado a la presencia de arcillas, sobre el contenido de potasio presente en otras alteraciones, además de los cationes intercambiables sodio y calcio.
- Al/Mg : Cuantifica enriquecimiento de aluminio asociado a la presencia de arcillas con respecto al magnesio, el cual se encuentra en mayor cantidad en alteraciones cloríticas y biotíticas.

- $3 * Al / (K + Na)$: También cuantifica el enriquecimiento de aluminio. La componente (K+Na) está asociada a una alteración sericítica.

Alteración argílica avanzada

- $(K + Al + S) / (Fe + S)$: Pretende destacar la alunita con respecto a la pirita (FeS_2)

Alteración potásica

- $K / (Ca + Na)$: Cuantifica el potasio por sobre el calcio y sodio.
- K / Mg : Resalta la sericita sobre la clorita.

Supérgena

- Mg / Al y Mn / Al : En las alteraciones supérgenas existe una mayor cantidad de magnesio y manganeso.

4.2.2. Tratamiento de variables geológicas

Previo a entrar de lleno al estudio exploratorio de datos, es importante analizar la cantidad de datos presente por cada tipo de litología, alteración y mineralización, con la finalidad de identificar la existencia de variables geológicas con pocos datos y/o que pertenezcan a unidades con poca relevancia para este estudio. La cantidad de datos por cada variable geológica se encuentra detallada en la Figura 4.1. Los códigos para cada litología, alteración y mineralización están detallados en las Tablas 4.2, 4.3 y 4.4.

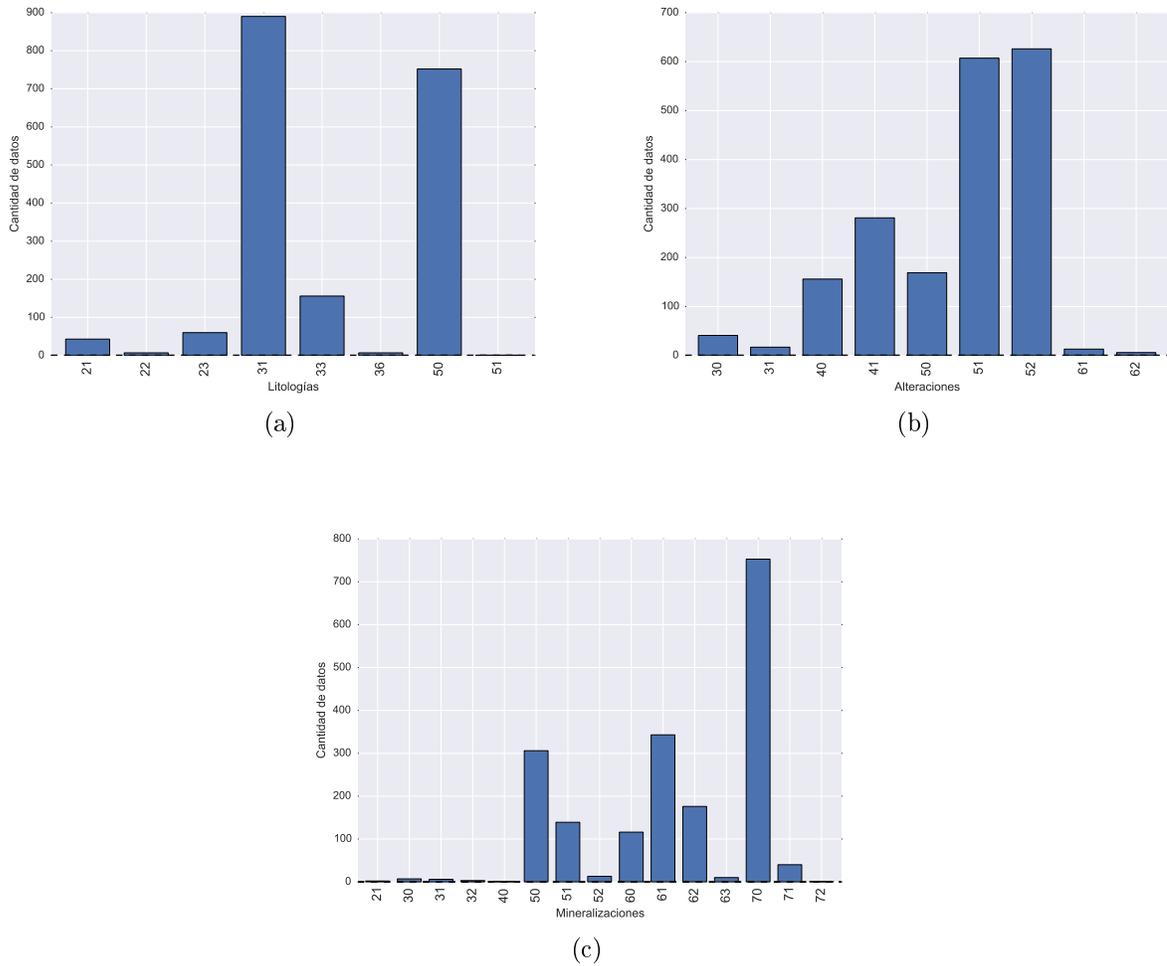


Figura 4.1: Cantidad de datos por cada tipo de variable geológica. (a): Litologías. (b): Alteraciones. (c): Mineralizaciones

Tabla 4.2: Códigos litologías

Litología	Código
Pórfido Feldespático	31
Pórfido Cuarzífero	33
Pórfido sin identificar	36
Andesita	50
Autobrecha (volcánica)	51
Brecha Hidrotermal	21
Brecha Tectónica	22
Brecha Magmática-Ígnea-Contacto	23

Tabla 4.3: Códigos alteraciones

Alteración	Código
Potásico (solo Bt secundaria)	30
Feld K >Bt	31
Clorita-Sericita-Arcillas	40
Clorita-Sericita-Cuarzo	41
Sericita (sin identificar)	50
Sericita-Cuarzo	51
Sericita-Cuarzo-Arcilla	52
Argilización Supérgena	61
Argilización Avanzada	62

Tabla 4.4: Códigos mineralizaciones

Mineralizaciones	Código
Óxidos Verdes (Sulfatos)	21
Parcial lixiviado desconocido	30
Parcial lixiviado (Oxidación parcial)	31
Parcial lixiviado (Limonitas exóticas)	32
Mezcla OxCu + Sulf	40
Cc + Py	50
Cc + Cv + Py	51
Cv + Py	52
Cc + Cpy + Py	60
Cc + Cv + Cpy + Py	61
Cv + Cpy + Py	62
Cv + Cpy + Bo (+/- Cc, +/- Py)	63
Py	72
Cpy - Py	70
Bn - Cpy (+/- Py)	71

Se observa en la Figura 4.1 que ciertas variables geológicas presentan baja cantidad de datos.

En las litologías, las brechas (códigos 21, 22 y 23), el Pórfido sin identificar (código 36) y la Autobrecha (código 51) tienen una baja cantidad de datos, por lo que serán omitidas.

En cuanto a las alteraciones, se omitirán las potásicas (códigos 30 y 31) y las argílicas supérgena y avanzada (códigos 61 y 62 respectivamente) por poseer una baja cantidad de datos. En cuanto al código 50, el cual es alteración Sericítica sin identificar, se debe destacar que pertenece a una codificación antigua, por lo que puede pertenecer tanto a las alteraciones Cuarzo-Sericita (51) como a la Cuarzo-Sericita-Arcilla (52). Considerando esto, se decidió no omitir este código e incluirlo en la base de datos final.

En las mineralizaciones, se decidió omitir los óxidos tanto de cobre como estériles (códigos 21, 30, 31, 32 y 40) junto con la pirita (código 72), debido a su bajísima cantidad de datos e irrelevancia en la recuperación de cobre por flotación rougher. Para facilitar la clasificación de las mineralizaciones se decidió agrupar códigos de acuerdo a la zona mineralizada a la que pertenecen. En el caso de los códigos 50, 51 y 52, pertenecientes a la zona de *alto enriquecimiento*, se les agrupó (re-nombró) al código HE. Los códigos 60, 61 y 62 son de la zona de *bajo enriquecimiento*, por lo tanto fueron renombrados a LE. Los códigos 70 y 71 son sulfuros primarios, por lo que serán catalogados bajo el nombre SP.

Finalmente, se re-nombraron el resto de los códigos de litologías y alteraciones para facilitar la lectura de los gráficos y resultados posteriores. El detalle de las nuevas codificaciones se explicitan en las tablas 4.5, 4.6 y 4.7.

Tabla 4.5: Nuevos códigos de litologías

Litologías		
Código antiguo	Código nuevo	Descripción
31	PF	Pórfido Feldespático
33	PC	Pórfido Cuarzífero
50	AN	Andesita

Tabla 4.6: Nuevos códigos de alteraciones

Alteraciones		
Código antiguo	Código nuevo	Descripción
40	CSA	Clorita-Sericita-Arcillas
41	CSC	Clorita-Sericita-Cuarzo
50	SSI	Sericita sin identificar
51	SCC	Sericita-Cuarzo
52	SCA	Sericita-Cuarzo-Arcillas

Tabla 4.7: Nuevos códigos de mineralizaciones

Mineralizaciones		
Código antiguo	Código nuevo	Descripción
50, 51 y 52	HE	Alto Enriquecimiento
60, 61, 62 y 63	LE	Bajo Enriquecimiento
70 y 71	SP	Sulfuros Primarios

4.3. Estudio exploratorio de datos

Una vez eliminadas las variables geológicas irrelevantes, se obtuvo la base de datos final con 1706 datos. La zona de estudio es de aproximadamente 2.44 kilómetros de ancho, 1.51 kilómetros de largo y 550 metros en el eje de profundidad (Tabla 4.8).

Tabla 4.8: Coordenadas de los datos.

Coordenada	Desde [m]	Hasta [m]	Diferencia [m]
Este	15398.95	17800.8	2401.85
Norte	105999.45	107513.89	1514.44
Cota	2440.98	2992.1	551.12

Se observa en la Figura 4.2 que los datos fueron muestreados exactamente al sur de mina Escondida.

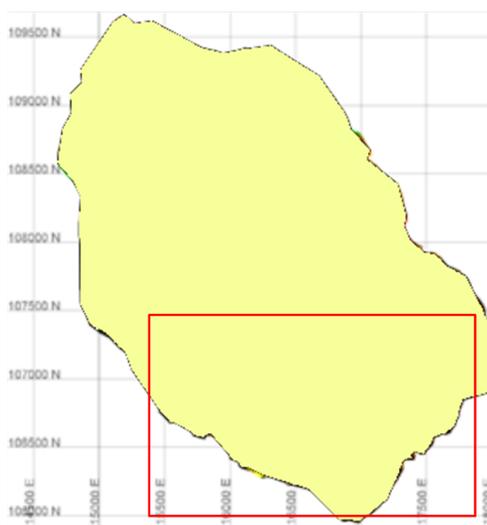


Figura 4.2: Vista básica en planta de mina Escondida. En el cuadrado se encuentra la zona donde se distribuyen los datos a analizar. Modificado de [1]

A continuación se detallarán las estadísticas básicas de los datos de geoquímica, variables sintéticas y recuperación para la base de datos final. Sin embargo, si se desea revisar este estudio para la base de datos original, es decir, con las variables geológicas originales, los detalles se encuentran en los Anexos A.1.

4.3.1. Datos de geoquímica

En cuanto a las estadísticas básicas, se observa en la Tabla 4.9 que prácticamente todas las variables presentan 1706 datos, salvo el oro (Au) que solo posee 871, lo cual es aproximadamente un 51% del total. Para efectos de este trabajo no se aplicará un algoritmo de imputación sobre los datos faltantes, por lo que se trabajará únicamente con los datos disponibles.

A su vez se puede apreciar que varios elementos presentan rangos intercuartiles² (RIC) nulos (Au, B, Hf, Nb, Ta, Ti y Zr), lo que indica la poca variabilidad de estas variables. Sin embargo, a priori no deben ser eliminadas debido a que sus valores extremos pueden presentar propiedades importantes en la predicción de la recuperación de cobre.

Además existe un alto coeficiente de variación³ ($CV > 3$) en 5 variables: As, Cd, Pb, Sb y W, lo que indica que estas variables presentan una alta variabilidad con respecto al resto.

Tabla 4.9: Estadísticas generales

Elemento	N° de datos	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
Ag	1706	3.48	3.39	0.97	0.04	1.48	2.70	4.62	51.30	3.14
Al	1706	0.82	0.59	0.71	0.08	0.42	0.60	0.99	3.81	0.57
As	1706	48.94	217.26	4.44	0.10	3.70	11.35	32.50	6610.00	28.80
Au	871	0.21	0.09	0.43	0.20	0.20	0.20	0.20	1.90	0.00
B	1706	10.18	1.48	0.15	10.00	10.00	10.00	10.00	40.00	0.00
Ba	1706	37.37	19.97	0.53	4.80	30.00	30.00	40.00	240.00	10.00
Be	1706	0.24	0.23	0.94	0.05	0.08	0.14	0.35	1.56	0.27
Bi	1706	1.31	1.65	1.26	0.03	0.34	0.81	1.66	19.85	1.32
Ca	1706	0.14	0.20	1.50	0.01	0.02	0.06	0.19	3.89	0.17
Cd	1706	4.19	13.35	3.18	0.01	0.31	1.17	3.53	345.00	3.22
Ce	1706	17.74	9.00	0.51	0.69	10.70	16.65	24.30	45.50	13.60
Co	1706	12.34	12.15	0.99	0.80	6.50	10.60	15.68	349.00	9.18
Cr	1706	115.45	111.37	0.96	1.00	37.25	83.00	164.00	1560.00	126.75
Cs	1706	1.06	0.55	0.52	0.15	0.68	0.92	1.32	4.60	0.64
Cu	1706	0.70	0.23	0.33	0.17	0.51	0.69	0.93	1.00	0.42
Fe	1706	2.29	1.32	0.58	0.00	1.49	2.00	2.76	13.58	1.27
Ga	1706	2.24	1.94	0.87	0.25	0.89	1.37	2.92	11.35	2.03
Ge	1706	0.06	0.03	0.42	0.05	0.05	0.05	0.07	0.58	0.02
Hf	1706	0.02	0.00	0.17	0.02	0.02	0.02	0.02	0.09	0.00
Hg	1706	0.03	0.07	2.31	0.01	0.01	0.01	0.03	1.44	0.02
In	1706	0.32	0.43	1.34	0.01	0.08	0.17	0.38	4.92	0.30
K	1706	0.21	0.07	0.35	0.01	0.16	0.20	0.24	0.76	0.08
La	1706	8.83	4.40	0.50	0.30	5.50	8.40	12.10	23.90	6.60
Li	1706	4.40	6.52	1.48	0.00	1.00	1.00	5.00	33.00	4.00
Mg	1706	0.35	0.53	1.54	0.01	0.03	0.04	0.45	2.39	0.42
Mn	1706	0.02	0.03	1.70	0.00	0.00	0.00	0.02	0.29	0.02
Mo	1706	93.09	95.25	1.02	2.79	36.23	63.90	118.30	1085.00	82.08
Na	1706	0.06	0.02	0.45	0.01	0.04	0.05	0.07	0.21	0.03
Nb	1706	0.05	0.01	0.26	0.05	0.05	0.05	0.05	0.23	0.00
Ni	1706	13.74	17.29	1.26	1.40	4.40	7.00	13.70	200.30	9.30
P	1706	381.95	448.98	1.18	10.00	70.00	160.00	537.50	2910.00	467.50
Pb	1706	56.14	176.36	3.14	0.30	4.20	19.50	56.88	4550.00	52.68
Rb	1706	11.98	5.33	0.44	0.80	8.90	10.80	13.40	55.40	4.50
Re	1706	0.72	1.23	1.71	0.00	0.15	0.36	0.72	20.70	0.57
S	1706	2.06	1.38	0.67	0.37	1.23	1.76	2.46	10.00	1.23
Sb	1706	1.97	10.73	5.45	0.05	0.13	0.25	0.83	306.00	0.70
Sc	1706	1.36	1.52	1.12	0.10	0.50	0.70	1.60	13.30	1.10
Se	1706	3.82	1.64	0.43	0.40	2.70	3.50	4.60	15.50	1.90
Sn	1706	0.75	1.76	2.35	0.20	0.30	0.50	0.70	59.00	0.40
Sr	1706	104.31	131.09	1.26	4.40	31.63	62.20	119.00	1240.00	87.38
Ta	1706	0.01	0.00	0.25	0.01	0.01	0.01	0.01	0.07	0.00
Te	1706	0.71	1.55	2.18	0.03	0.21	0.37	0.65	26.70	0.44
Th	1706	1.79	0.98	0.55	0.20	1.00	1.60	2.50	6.10	1.50
Ti	1706	0.01	0.02	1.72	0.01	0.01	0.01	0.01	0.18	0.00
Tl	1706	0.17	0.11	0.62	0.02	0.11	0.15	0.20	1.03	0.09
U	1706	0.44	0.31	0.70	0.05	0.26	0.36	0.51	3.28	0.25
V	1706	13.76	19.06	1.38	1.00	2.00	5.00	15.00	168.00	13.00
W	1706	0.57	1.91	3.37	0.05	0.18	0.32	0.55	50.10	0.37
Y	1706	4.09	5.50	1.35	0.06	0.50	0.97	6.83	48.00	6.33
Zn	1706	495.97	782.97	1.58	2.00	24.00	182.00	718.75	10000.00	694.75
Zr	1706	1.00	0.02	0.02	1.00	1.00	1.00	1.00	2.00	0.00

²Rango intercuartil (RIC), es la diferencia entre el cuartil 3 (75 %) y el cuartil 1 (25 %)

³Coefficiente de variación o desviación relativa (CV), corresponde a la razón entre la desviación estándar y la media

Tabla 4.9: Estadísticas generales

Elemento	N° de datos	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
----------	-------------	-------	------------	----	--------	------	------	------	--------	-----

4.3.2. Variables sintéticas

Las estadísticas asociadas a las variables sintéticas definidas se encuentran en la Tabla 4.10. A priori no se pueden sacar conclusiones con respecto a estas estadísticas, debido a que muchas de estas variables están asociadas a diferentes alteraciones, por lo que deben ser analizadas una vez definidos los dominios.

Tabla 4.10: Estadísticas de las variables sintéticas

Variable	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
cu/s	0.45	0.27	0.60	0.04	0.24	0.39	0.60	1.61	0.36
k/al	0.34	0.14	0.43	0.05	0.22	0.33	0.45	0.70	0.23
kxal	0.20	0.21	1.05	0.00	0.08	0.12	0.23	1.92	0.16
(kxna)/al	0.02	0.01	0.40	0.00	0.01	0.02	0.02	0.10	0.01
na/al	0.08	0.04	0.50	0.01	0.06	0.08	0.10	1.00	0.05
(al+k)/(na+ca+mg)	4.37	3.16	0.72	0.13	1.48	3.64	6.70	16.33	5.22
(al+k+na)/(ca+mg)	7.69	6.83	0.89	0.13	1.71	5.00	12.80	37.50	11.09
(ca+na)/(k+al)	0.20	0.28	1.35	0.03	0.10	0.14	0.23	7.80	0.14
al/(na+ca+k)	1.89	1.31	0.70	0.05	0.86	1.64	2.62	10.50	1.76
al/mg	11.84	9.29	0.79	0.24	1.94	12.50	18.33	48.00	16.40
3*al/(na+k)	8.77	4.75	0.54	1.85	5.55	7.13	10.26	35.67	4.71
(k+al+s)/(fe+s)	0.75	0.14	0.18	0.34	0.66	0.73	0.82	1.91	0.16
k/mg	4.66	4.22	0.91	0.09	0.43	4.60	7.25	27.00	6.82
mg/al	0.27	0.31	1.15	0.02	0.05	0.08	0.52	4.19	0.46
mn/al	0.02	0.02	1.38	0.00	0.00	0.01	0.02	0.32	0.02

La matriz de correlación entre todas las variables presentes en la base de datos, incluyendo las variables sintéticas, se puede visualizar en la Figura 4.3, de la cual se puede observar que ciertas variables tales como Au, Hf, Zr, Ta, W y B presentan muy baja correlación con respecto al resto de las variables, lo cual indicaría que se deben eliminar del estudio. Sin embargo, el que estos elementos no estén correlacionados con otros no implica que no lo estén con respecto a la recuperación, por lo que requiere un análisis en ese sentido.

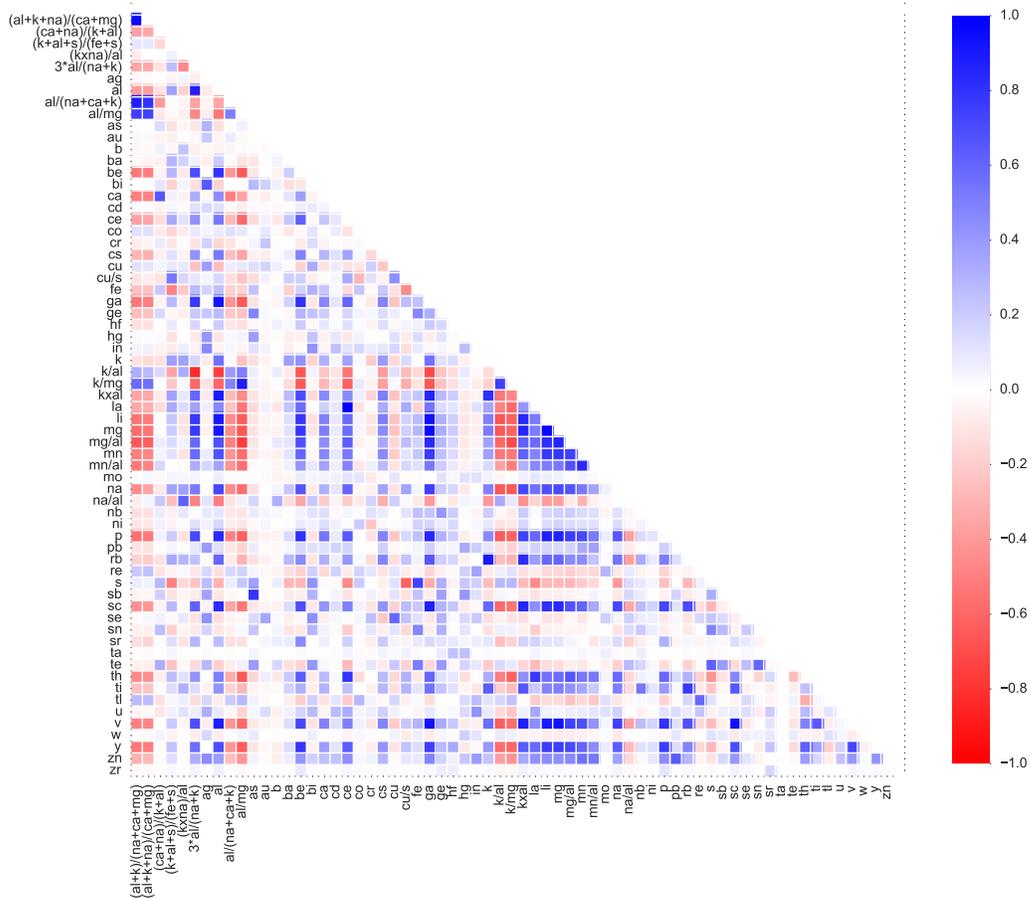


Figura 4.3: Matriz de correlación con todas las variables presentes en el estudio, incluyendo variables sintéticas.

4.3.3. Datos de recuperación

En cuanto a las estadísticas de la recuperación de cobre (Tabla 4.11), es importante destacar el bajo coeficiente de variación de los datos (0.09) y un rango intercuartil de menos de un 9% en el total de los datos, por lo que es relevante estudiar las estadísticas de la recuperación por cada dominio a definir.

Tabla 4.11: Estadísticas de recuperación

Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
84.88 %	7.49 %	0.09	36.47 %	81.40 %	86.69 %	90.27 %	98.14 %	8.87 %

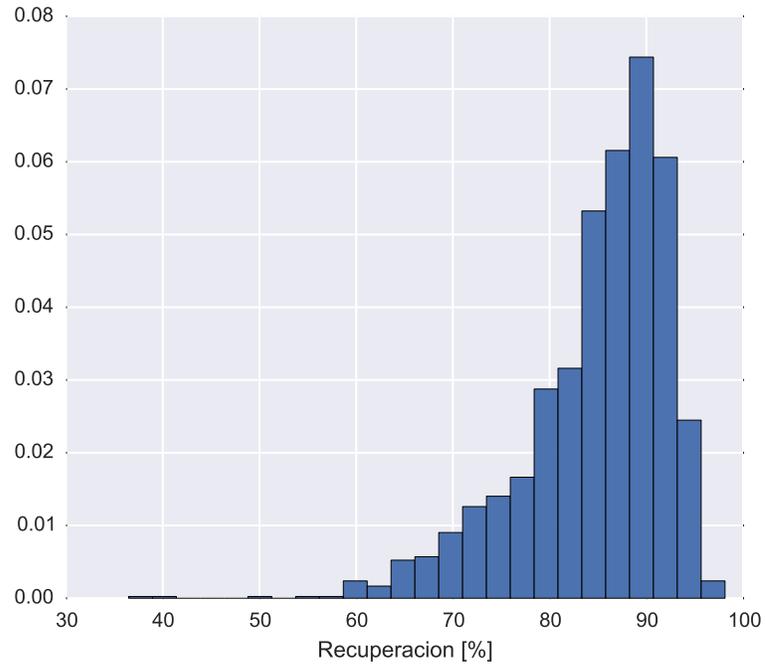
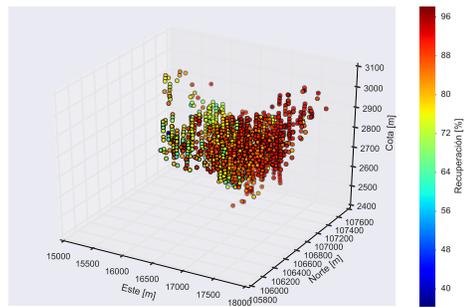
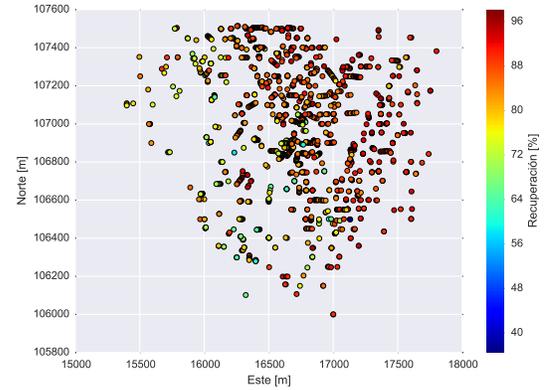


Figura 4.4: Histograma de recuperación de cobre.

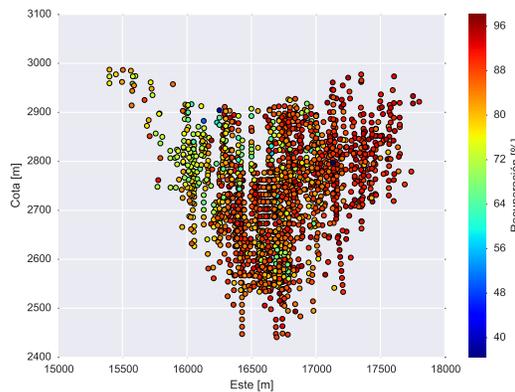
Se observa en la Figura 4.5 la distribución espacial de la recuperación en vista isométrica y en cada uno de los tres planos (Este-Norte, Este-Cota y Norte-Cota).



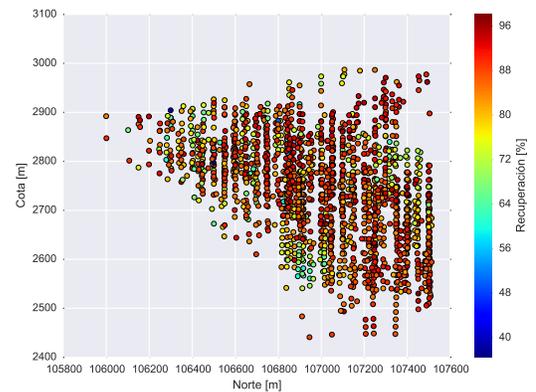
(a)



(b)



(c)



(d)

Figura 4.5: Vista espacial de la recuperación de cobre. (a): Vista isométrica de sondajes con sus valores de recuperación de cobre. (b): Vista en el plano Este-Norte. (c): Vista en el plano Este-Cota. (d): Vista en el plano Norte-Cota.

4.3.4. Datos de geología

Con respecto a las litologías (Figura 4.6), se observa que los pórfidos cuarcífero y feldespático presentan una recuperación de cobre mayor que la andesita.

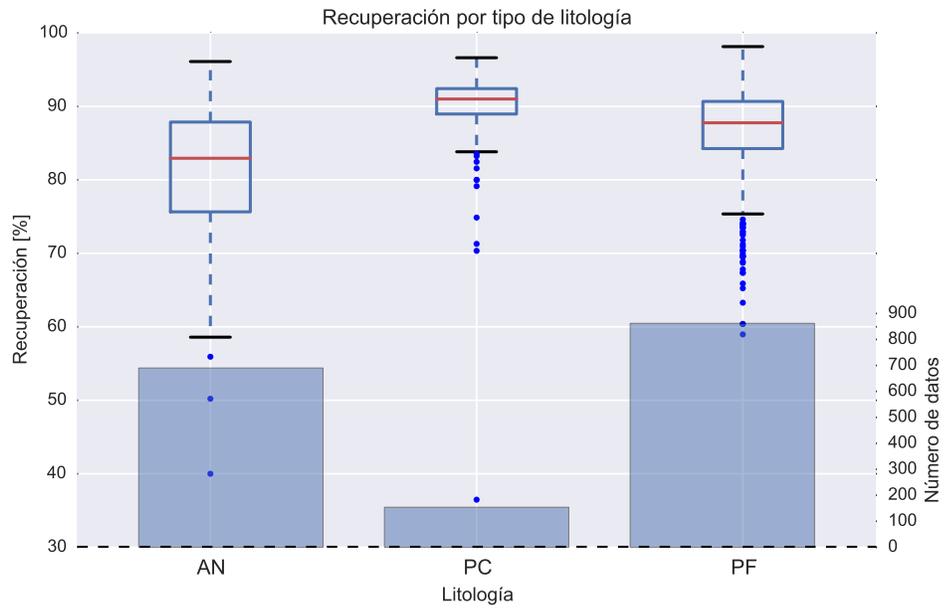


Figura 4.6: Recuperación por tipo de litología

Con respecto a las alteraciones (Figura 4.7), se observa que la recuperación media es levemente inferior en alteraciones con presencia de arcillas (Clorita-Sericita-Arcillas (CSA) y Sericita-Cuarzo-Arcillas (SCA)).

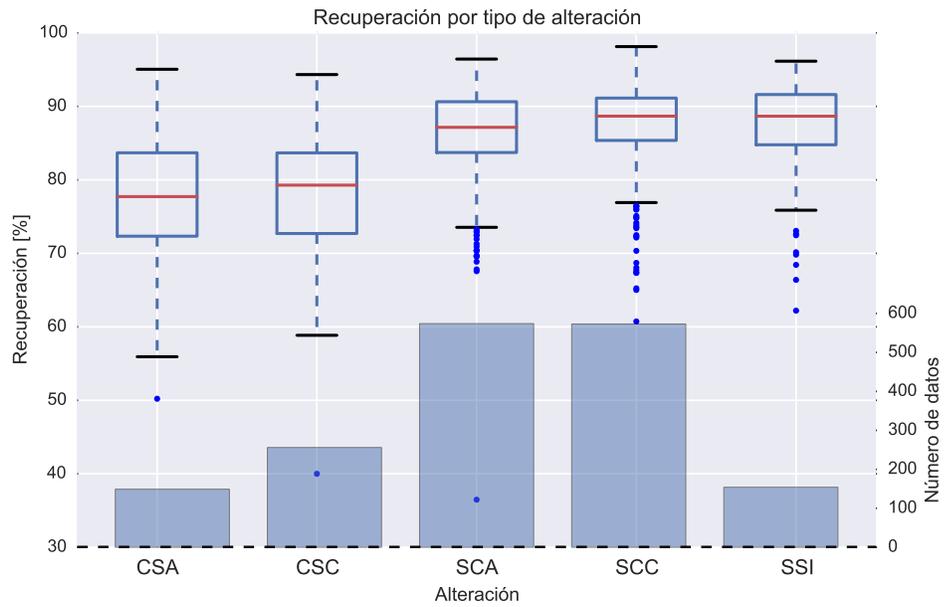


Figura 4.7: Recuperación por tipo de alteración

En cuanto a las mineralizaciones (Figura 4.8), la zona de alto enriquecimiento (HE) presenta una mayor recuperación que la de bajo enriquecimiento (LE) y la de sulfuros primarios (SP), permitiendo inferir una relación directa entre el contenido de cobre en la mineralización y su recuperación.

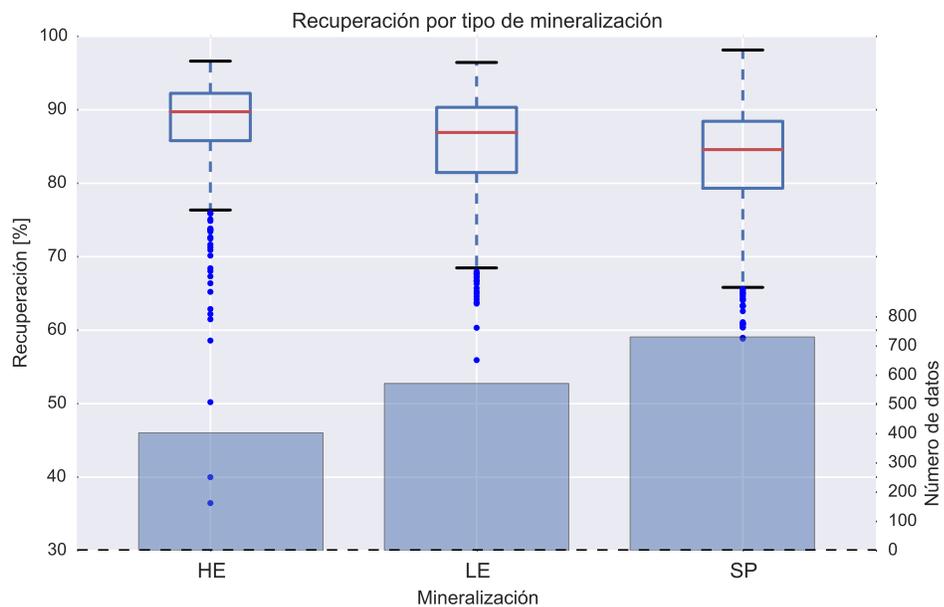


Figura 4.8: Recuperación por tipo de mineralización

Capítulo 5

Definición de dominios geometalúrgicos

En este capítulo se detallará la aplicación de la metodología propuesta para la definición de dominios geometalúrgicos, consistente en:

1. Realizar una selección de variables relevantes, es decir, que presenten una mayor correlación con la recuperación de cobre y no sean redundantes con otras variables dentro del estudio.
2. Normalizar los datos de las variables seleccionadas y aplicar análisis de componentes principales sobre ellas.
3. Seleccionar las primeras componentes principales en función de su varianza total explicada.
4. Aplicar y comparar métodos de clustering realizados sobre las componentes principales. En este paso es cuando se definen los dominios a estudiar.
5. Seleccionar un método de clustering en base a criterios definidos y analizar a fondo los clusters o dominios obtenidos.
6. Realizar modelos predictivos (regresión multilínea) para cada uno de los dominios definidos.
7. Detallar un análisis de sensibilidad, donde se compare la calidad de agrupación de diversas variables geológicas en función de la cantidad de dominios o clusters definidos.

Esta propuesta de definición de modelos, en base a minería de datos, es diferente a la metodología clásica, consistente en la definición de dominios en función solamente de características geológicas. En el Anexo C se aplica y detalla este enfoque tradicional sobre los datos utilizados en este trabajo.

5.1. Selección de variables

La definición de los dominios geometalúrgicos será determinada en base a 66 variables, de las cuales 51 son de geoquímica y 15 son sintéticas.

Al realizar un análisis de componentes principales sobre esta gran cantidad de variables, se corre el riesgo de que las primeras componentes principales expliquen una baja cantidad de varianza total, debido a una gran cantidad de variables que generan ruido en los resultados, por lo que es necesario omitir variables irrelevantes y/o redundantes previo a la definición de los dominios.

Tabla 5.1: Correlaciones entre recuperación y variables geoquímicas y sintéticas. Ordenados de mayor a menor.

K/Al	0.58	(Ca+Na)/(K+Al)	0.06	Mn/Al	-0.28
K/Mg	0.51	W	0.05	Ca	-0.29
Al/Mg	0.50	In	0.03	K	-0.30
(Al+K)/(Na+Ca+Mg)	0.39	Ta	0.01	Rb	-0.32
(Al+K+Na)/(Ca+Mg)	0.35	Nb	0.00	Cs	-0.42
Al/(Na+Ca+K)	0.34	Zr	0.00	La	-0.44
Cu	0.33	Ni	-0.02	Th	-0.48
S	0.25	Cu/S	-0.03	Ce	-0.48
Na/Al	0.24	Co	-0.04	Mn	-0.49
Bi	0.20	Ba	-0.07	Sc	-0.54
Ag	0.18	B	-0.07	Na	-0.55
Se	0.16	Cd	-0.07	KxAl	-0.57
Te	0.15	Pb	-0.07	Y	-0.58
Cr	0.14	Hf	-0.11	Mg/Al	-0.58
As	0.14	U	-0.12	V	-0.60
Tl	0.13	(K+Al+S)/(Fe+S)	-0.15	Ga	-0.63
Sb	0.10	Ge	-0.16	P	-0.65
Sn	0.10	Mo	-0.16	Mg	-0.66
(KxNa)/al	0.09	Fe	-0.19	Be	-0.66
Au	0.09	Ti	-0.19	3*Al/(Na+K)	-0.67
Hg	0.07	Zn	-0.26	Li	-0.69
Re	0.07	Sr	-0.26	Al	-0.70

Las variables con baja correlación tanto positiva o negativa pueden ser omitidas del estudio, permitiendo que solo variables con alta correlación sean consideradas para la definición de dominios y su posterior modelo predictivo multilíneal. Sin embargo, se requiere un criterio para definir un intervalo en el cual sea pertinente eliminar variables no relevantes y un valor mínimo de correlación para determinar si una variable es redundante con respecto a otra.

El criterio escogido para la selección de variables relevantes es que posean una correlación positiva mayor a 0.25 y/o una negativa inferior a -0.6. Estos valores fueron escogidos en base a los resultados de la Tabla 5.1, por lo tanto, dependiendo del criterio a utilizar, se pueden utilizar otros valores en la selección. En cuanto al criterio de redundancia entre variables, a partir del trabajo de Boisvert et al. [2] se determina que la correlación mínima para asegurar que una variable es redundante con respecto a otra es mayor o igual a 0.94.

Aplicando el primer criterio sobre los datos presentes en la Tabla 5.1, se obtuvo que las variables seleccionadas por relevancia son: K/Al , K/Mg , Al/Mg , $(Al+K)/(Na+Ca+Mg)$,

$(Al+K+Na)/(Ca+Mg)$, $Al/(Na+Ca+K)$, Cu y S , V , Ga , P , Mg , Be , $3*Al/(Na+K)$, Li y Al . Posteriormente se aplicó el criterio de redundancia en base a las correlaciones entre elementos (Anexo A.2.1), donde se decidió eliminar las siguientes variables:

- $(Al + K + Na)/(Ca + Mg)$: ya que presenta correlación de 0.96 con la variable $(Al + K)/(Na + Ca + Mg)$ y esta última variable presenta mejor correlación con respecto a la recuperación de cobre (0.39).
- Ga : Presenta una alta correlación (0.96) con Mg . Se elimina Ga por presentar menor concentración.
- Li : Presenta una alta correlación (0.96) con Mg . Se elimina Li por presentar menor concentración.

Finalmente se redujo el modelo de 66 a 13 variables, las cuales son:

- K/Al
- K/Mg
- Al/Mg
- $(Al + K)/(Na + Ca + Mg)$
- $Al/(Na + Ca + K)$
- Cu
- S
- V
- P
- Mg
- Be
- $3 * Al/(Na + K)$
- Al

5.2. Análisis de componentes principales

Una vez seleccionadas las variables relevantes en el estudio, se realizó sobre ellas un análisis de componentes principales (PCA). El objetivo del PCA es reducir la dimensionalidad de los datos, a través de la búsqueda de combinaciones lineales de las variables originales que representen lo mejor posible la variabilidad presente en los datos. Previo a la implementación de PCA es necesario normalizar los datos para evitar que variables más abundantes oculten la variabilidad de variables menos abundantes. La cantidad original de componentes principales es igual al número de variables del modelo (Tablas 5.2 y 5.3). De igual manera, se deben elegir una cantidad limitada de estos para permitir su visualización y posterior implementación de métodos de clustering aplicado sobre estas componentes.

En base a los resultados de la Tabla 5.2 se observa que las dos primeras componentes principales explican el 92.04% de la varianza total y las tres primeras un 95.78%. Para ver cuantas componentes se utilizarán se hará uso del gráfico de sedimentación (Figura 5.1), en

este gráfico se observa un «codo» entre la segunda y la tercera componente, por lo que se decidió seleccionar las tres primeras componentes.

Tabla 5.2: Componentes principales junto con sus varianzas explicadas y acumulada. Componentes 1 a 6

	CP1	CP2	CP3	CP4	CP5	CP6
(Al+K)/(Na+Ca+Mg)	-0.227	0.289	-0.648	0.317	-0.227	0.228
3*Al/(Na+K)	-0.218	0.502	-0.074	-0.375	0.132	0.05
Al	-0.014	0.035	-0.02	0.004	0.01	-0.022
Al/(Na+Ca+K)	-0.088	0.092	-0.276	0.235	-0.321	0.141
Al/Mg	-0.802	-0.132	-0.109	-0.24	0.3	-0.242
Be	-0.002	0.005	-0.001	0	-0.002	-0.002
Cu	-0.029	0.04	-0.002	-0.049	-0.06	0.019
K/Al	-0.018	-0.01	0.007	-0.008	-0.065	-0.004
K/Mg	-0.358	-0.448	0.172	0.229	-0.537	-0.246
Mg	0.001	0.008	0.005	0.007	-0.002	-0.005
P	0.329	-0.112	-0.518	-0.446	-0.231	-0.598
S	-0.103	0.126	0.258	-0.567	-0.593	0.362
V	-0.049	0.638	0.349	0.264	-0.191	-0.563
Varianza Explicada	87.13 %	4.91 %	3.74 %	1.52 %	1.26 %	0.96 %
Varianza Acumulada	87.13 %	92.04 %	95.78 %	97.30 %	98.57 %	99.52 %

Tabla 5.3: Componentes principales junto con sus varianzas explicadas y acumulada. Componentes 6 a 13

	CP7	CP8	CP9	CP10	CP11	CP12	CP13
(Al+K)/(Na+Ca+Mg)	-0.153	-0.352	0.303	0.029	0	0.022	-0.009
3*Al/(Na+K)	0.717	0.074	0.097	0.054	-0.041	-0.008	0.003
Al	0.047	-0.036	0.023	-0.972	0.069	-0.16	0.138
Al/(Na+Ca+K)	0.055	0.668	-0.527	-0.022	-0.006	-0.033	0.014
Al/Mg	-0.318	0.085	-0.112	-0.003	0.028	0.006	-0.001
Be	0.015	0.008	-0.01	-0.145	0.035	0.046	-0.988
Cu	0.102	-0.623	-0.745	-0.004	-0.194	0.003	-0.002
K/Al	0.092	-0.123	-0.132	0.067	0.972	0.067	0.03
K/Mg	0.44	-0.099	0.175	0.011	-0.078	-0.007	-0.001
Mg	0.016	0.034	-0.007	-0.156	-0.057	0.983	0.067
P	-0.027	0.018	-0.002	0.008	-0.009	0.003	-0.001
S	-0.312	-0.007	0.069	-0.03	-0.005	0.003	-0.001
V	-0.211	-0.005	-0.009	0.019	0.007	-0.005	-0.001
Varianza Explicada	0.42 %	0.03 %	0.02 %	0.00 %	0.00 %	0.00 %	0.00 %
Varianza Acumulada	99.95 %	99.97 %	99.99 %	100.00 %	100.00 %	100.00 %	100.00 %

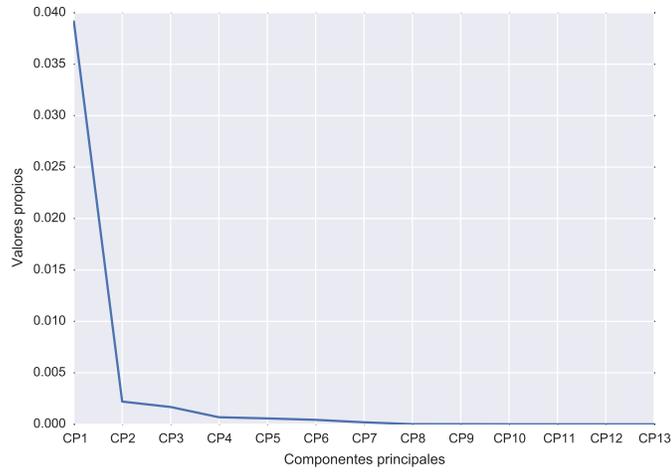


Figura 5.1: Gráfico de sedimentación

En las siguientes figuras se observan gráficamente las tres primeras componentes seleccionadas, tanto en vistas isométricas como en diferentes planos.

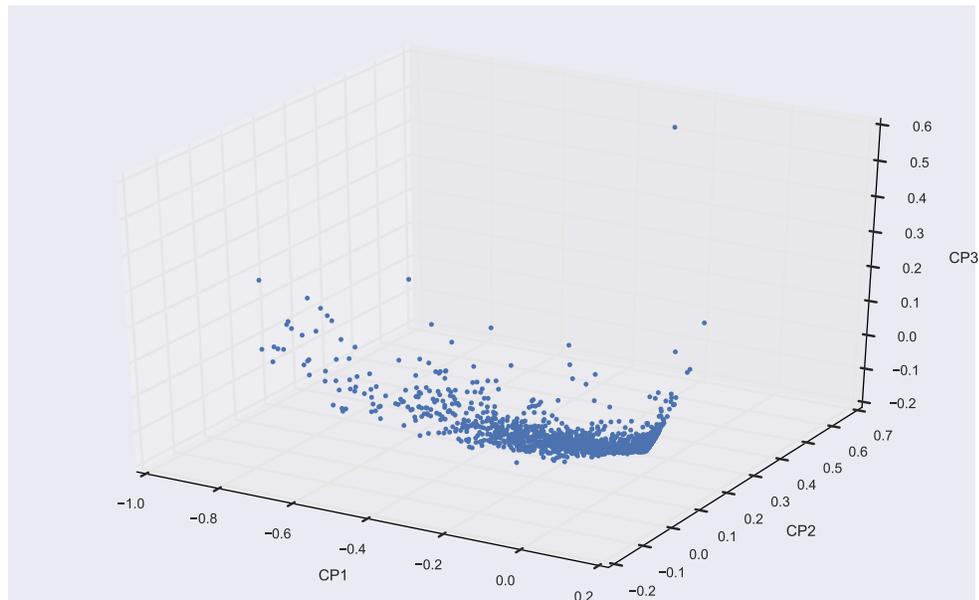


Figura 5.2: Vista isométrica de las tres componentes principales.

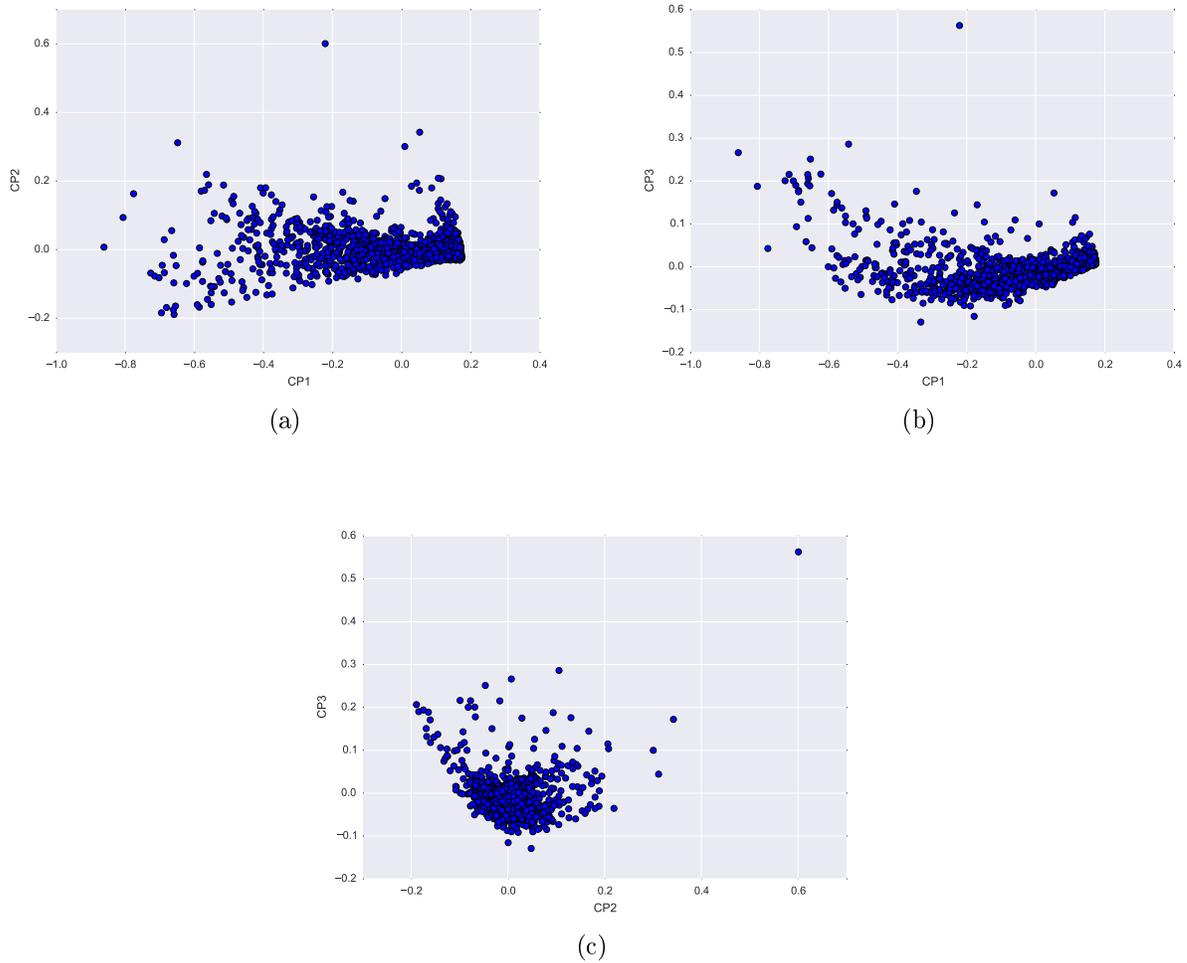


Figura 5.3: Visualización de resultados de PCA. (a): Vista bidimensional componentes 1 y 2. (b): Vista bidimensional componentes 1 y 3. (c): Vista bidimensional componentes 2 y 3.

Se puede observar a partir de los coeficientes obtenidos para la primera componente principal (Tabla 5.2), la cual además presenta la mayor cantidad de varianza explicada (87.13 %), que a medida que disminuye el valor de esta, aumenta fuertemente la razón Al/Mg , que cuantifica el enriquecimiento de aluminio asociado a la presencia de arcillas [21].

En base a lo expuesto anteriormente, en la Figura 5.4 se ilustra cómo la presencia de arcillas es mayor a menor valor de la primera componente principal.

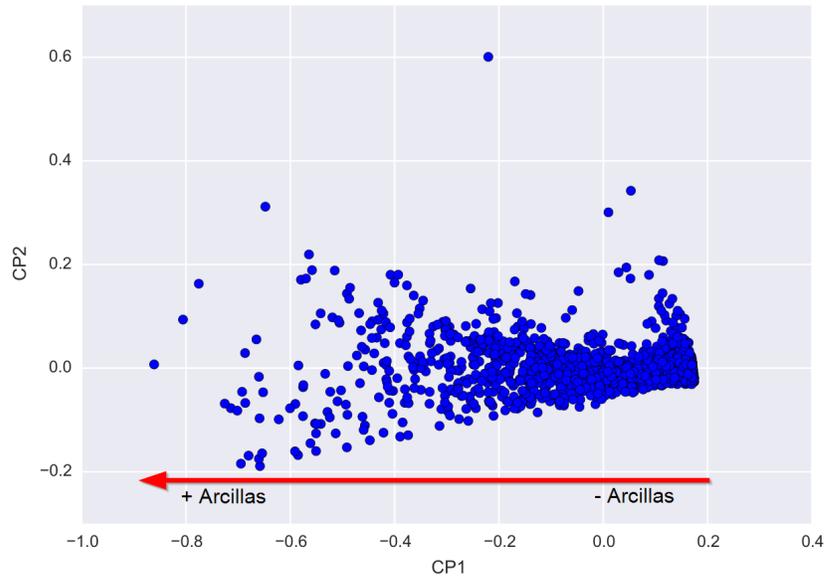


Figura 5.4: Presencia de arcillas en función de la primera componente principal

En cuanto a la segunda componente principal, si bien explica un bajo porcentaje de la varianza total (4.91 %), valores positivos de ella indican enriquecimiento de aluminio (coeficiente positivo de $3 * Al / (Na + K)$), lo cual permite sugerir un aumento de las arcillas en esa dirección (Figura 5.5).

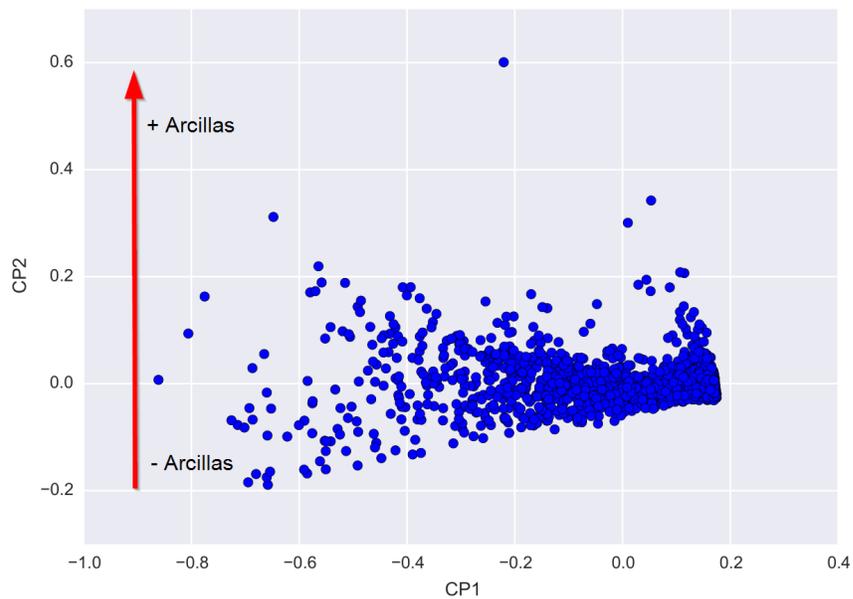


Figura 5.5: Presencia de arcillas en función de la segunda componente principal

Finalmente, con respecto a la tercera componente principal (varianza total explicada: 3.74 %), a menor valor de ésta se tiene mayor contenido de aluminio (coeficientes negativos

de $(Al + K)/(Na + Ca + Mg)$ y $Al/(Na + Ca + K)$). Una visualización de esta afirmación se observa en la Figura 5.6.

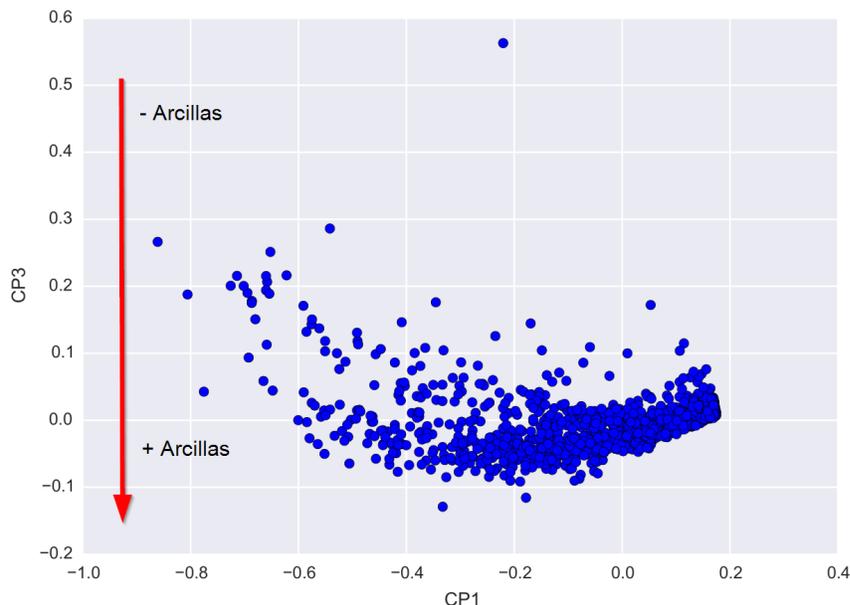


Figura 5.6: Presencia de arcillas en función de la tercera componente principal

5.3. Clustering

El agrupamiento o clustering se define como un procedimiento que permite dividir n muestras en k grupos. El objetivo del agrupamiento es maximizar la similitud entre las muestras de un mismo grupo y minimizar el parecido entre los datos de grupos distintos.

En este estudio se han implementado y comparado dos algoritmos diferentes de agrupamiento: k-Means y Ward, siendo el primero un método de particionamiento y el otro del tipo jerárquico aglomerativo.

Los clusters definidos por estos métodos serán considerados a priori como los dominios geometalúrgicos. Sin embargo, es relevante analizar cuan bien estos dominios clasifican geológicamente los datos y si es posible realizar un modelo predictivo de recuperación de cobre sobre los datos contenidos en ellos.

Estos dos algoritmos no calculan automáticamente la cantidad de clusters, por lo que es necesario utilizar una métrica o interpretar los resultados para cada conjunto de dominios definidos. En este caso se utilizó una métrica llamada *coeficiente de silueta*, que a mayor valor indica una mejor definición de los clusters. En la Figura 5.7 se calculó el coeficiente de silueta desde 2 a 25 clusters definidos para cada algoritmo.

Se observa que los mayores valores se obtienen con menos de 4 clusters, lo cual puede indicar que una baja cantidad de dominios es suficiente para clasificar los datos. Sin embargo, es necesario visualizar e interpretar los resultados obtenidos para comprobar esta afirmación.

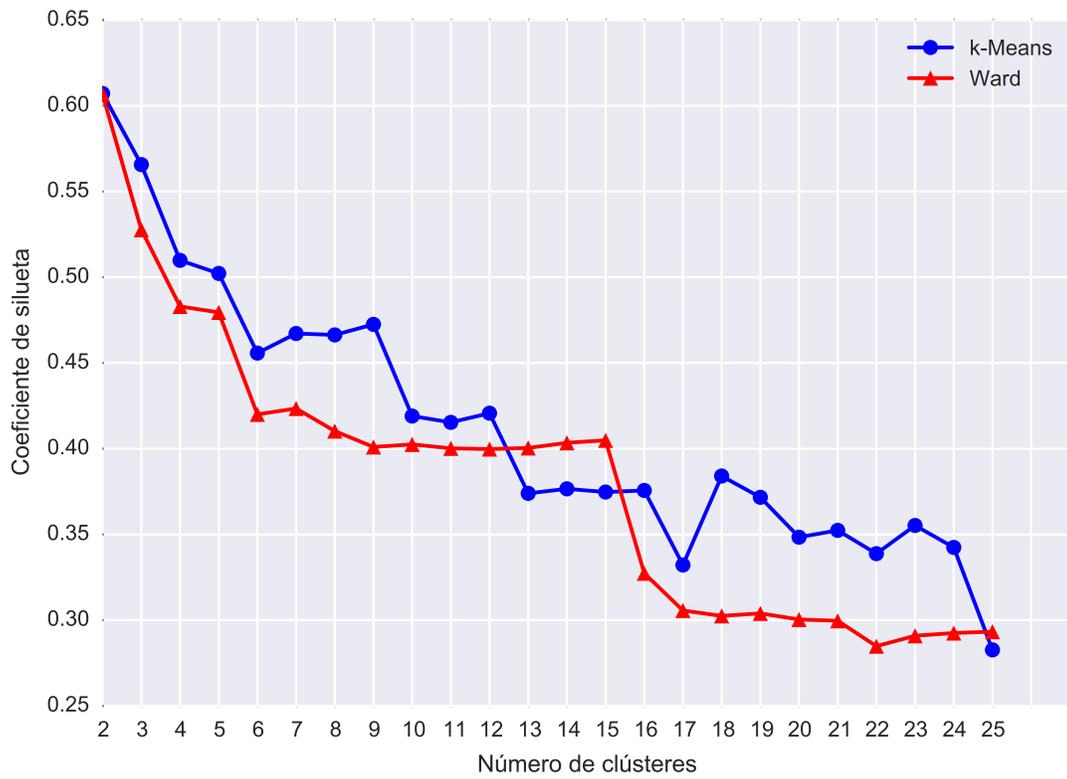


Figura 5.7: Coeficiente de silueta por cada algoritmo de clustering.

A continuación se mostrarán los resultados obtenidos al definir 4 dominios tanto por k-Means como a través de Ward.

5.3.1. k-Means

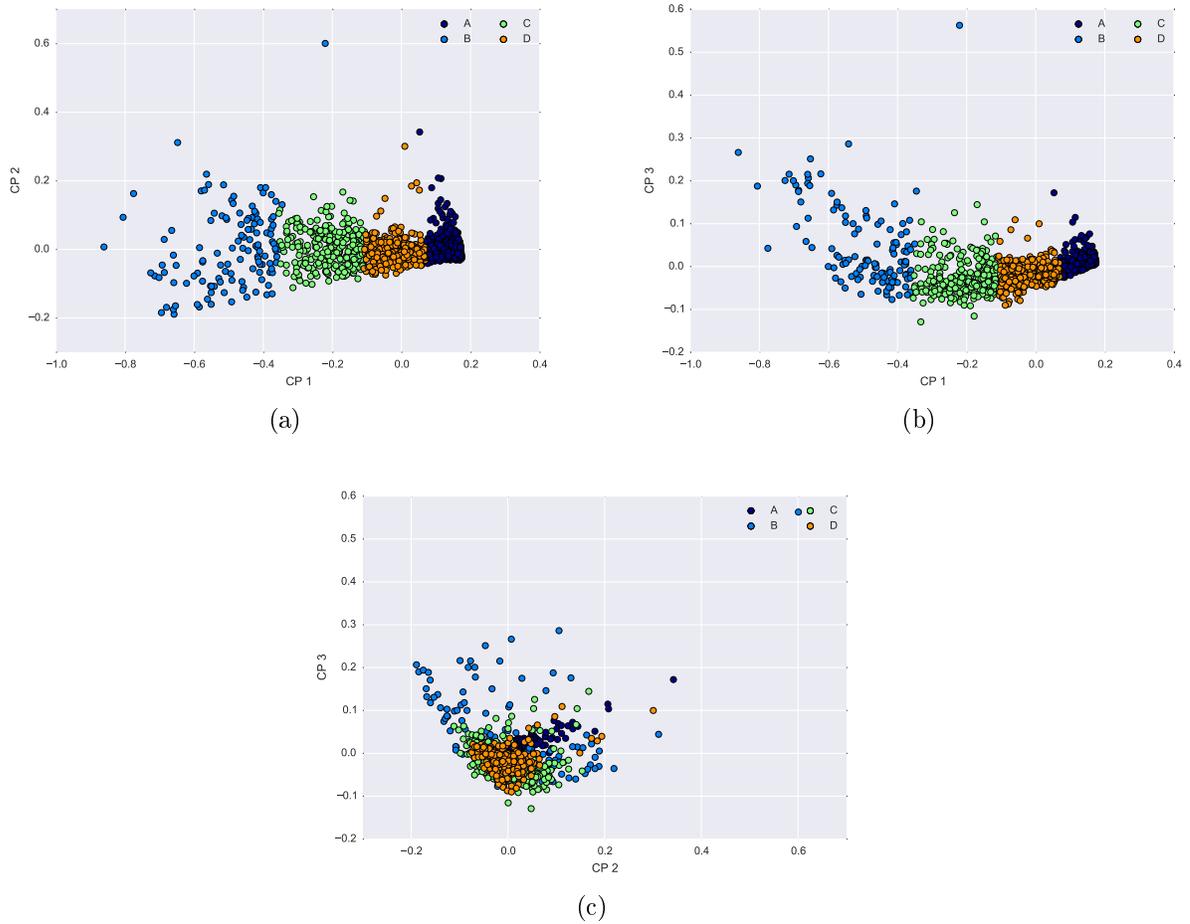


Figura 5.8: Resultados de clustering (k-Means de 4 dominios) sobre datos de PCA. (a): Vista bidimensional componentes 1 y 2. (b): Vista bidimensional componentes 1 y 3. (c): Vista bidimensional componentes 2 y 3.

A continuación se mostrarán estadísticas básicas de cada uno de los clusters definidos a través de k-Means.

5.3.1.1. Cluster A

Este cluster contiene 864 datos y sus estadísticas básicas se detallan en la Tabla 5.4.

Tabla 5.4: Estadísticas cluster A de k-Means

	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
$(Al+K)/(Na+Ca+Mg)$	2.50	1.86	0.74	0.13	1.18	1.65	3.24	11.38	2.06
$3*Al/(Na+K)$	11.16	5.41	0.48	1.85	7.06	9.52	13.82	35.67	6.75
Al	1.14	0.65	0.58	0.15	0.59	0.92	1.63	3.81	1.04
$Al/(Na+Ca+K)$	1.29	0.79	0.61	0.05	0.67	1.07	1.75	5.00	1.08
Al/Mg	5.08	5.96	1.17	0.24	1.37	1.96	6.47	32.00	5.10
Be	0.40	0.24	0.60	0.05	0.23	0.35	0.52	1.56	0.29
Cu	0.68	0.22	0.33	0.17	0.50	0.66	0.87	1.00	0.37
K/Al	0.25	0.11	0.44	0.05	0.16	0.24	0.33	0.63	0.16
K/Mg	1.60	2.19	1.37	0.09	0.23	0.45	2.11	12.00	1.89
Mg	0.66	0.60	0.92	0.01	0.10	0.43	1.20	2.39	1.10
P	674.91	471.21	0.70	60.00	250.00	530.00	1090.00	2910.00	840.00
S	1.75	1.17	0.67	0.37	1.04	1.43	2.09	10.00	1.05
V	24.04	22.31	0.93	1.00	6.00	15.00	38.00	168.00	32.00

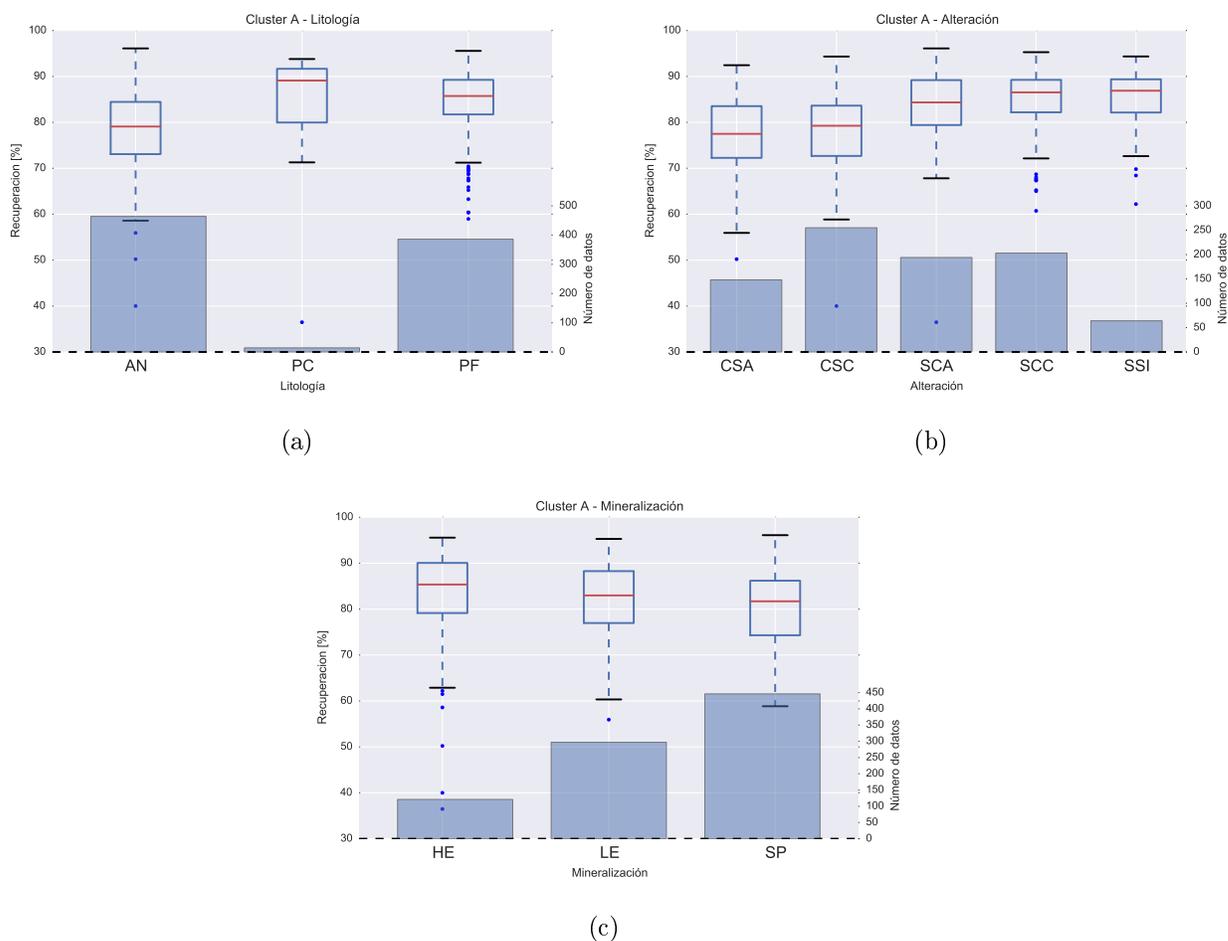


Figura 5.9: Boxplots de recuperación por tipo de variable geológica en el cluster A definido por k-Means. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.

5.3.1.2. Cluster B

Este cluster contiene 125 datos y sus estadísticas básicas se detallan en la Tabla 5.5.

Tabla 5.5: Estadísticas cluster B de k-Means

	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
$(Al+K)/(Na+Ca+Mg)$	5.50	3.30	0.60	0.29	2.33	5.67	7.56	14.40	5.22
$3*Al/(Na+K)$	6.05	2.06	0.34	3.75	4.43	5.25	7.11	13.29	2.68
Al	0.41	0.19	0.46	0.10	0.27	0.39	0.52	1.19	0.25
$Al/(Na+Ca+K)$	2.11	1.42	0.67	0.05	0.89	1.88	3.00	6.00	2.11
Al/Mg	21.20	7.21	0.34	5.67	16.50	21.00	24.00	48.00	7.50
Be	0.07	0.02	0.33	0.05	0.05	0.05	0.08	0.16	0.03
Cu	0.75	0.25	0.33	0.23	0.49	0.84	1.00	1.00	0.51
K/Al	0.45	0.15	0.32	0.15	0.31	0.47	0.58	0.70	0.27
K/Mg	9.68	5.06	0.52	1.67	5.50	9.00	12.50	27.00	7.00
Mg	0.02	0.01	0.47	0.01	0.01	0.02	0.03	0.04	0.02
P	30.80	10.36	0.34	10.00	20.00	30.00	40.00	60.00	20.00
S	2.66	1.83	0.69	1.08	1.67	2.08	2.96	10.00	1.29
V	3.02	2.96	0.98	1.00	2.00	2.00	4.00	30.00	2.00

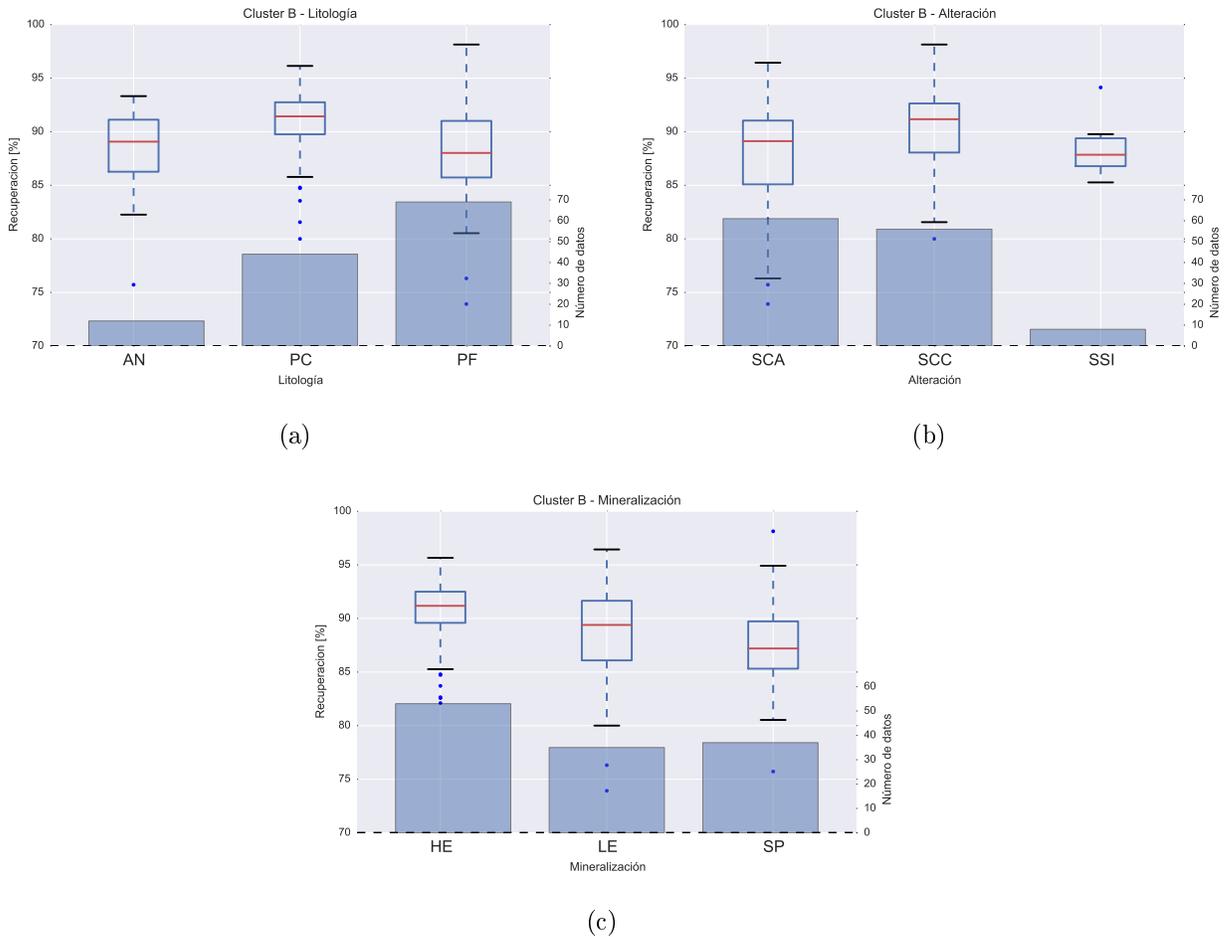


Figura 5.10: Boxplots de recuperación por tipo de variable geológica en el cluster B definido por k-Means. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.

5.3.1.3. Cluster C

Este cluster contiene 311 datos y sus estadísticas básicas se detallan en la Tabla 5.6.

Tabla 5.6: Estadísticas cluster C de k-Means

	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
$(Al+K)/(Na+Ca+Mg)$	6.82	3.13	0.46	0.50	4.80	7.00	8.88	16.00	4.09
$3*Al/(Na+K)$	6.51	2.27	0.35	3.63	4.98	5.69	7.73	19.32	2.75
Al	0.50	0.22	0.43	0.08	0.35	0.45	0.63	1.61	0.28
$Al/(Na+Ca+K)$	2.65	1.39	0.53	0.05	1.76	2.60	3.50	10.00	1.74
Al/Mg	19.51	6.30	0.32	8.00	15.00	18.00	23.00	47.00	8.00
Be	0.09	0.04	0.44	0.05	0.05	0.08	0.11	0.29	0.06
Cu	0.74	0.23	0.31	0.29	0.54	0.77	1.00	1.00	0.46
K/Al	0.41	0.13	0.31	0.11	0.31	0.43	0.51	0.70	0.20
K/Mg	7.84	3.08	0.39	1.00	5.67	7.00	9.50	19.00	3.83
Mg	0.03	0.01	0.35	0.01	0.02	0.03	0.03	0.05	0.01
P	59.32	19.61	0.33	30.00	40.00	60.00	70.00	150.00	30.00
S	2.29	1.44	0.63	0.73	1.48	1.91	2.61	10.00	1.13
V	2.97	1.69	0.57	1.00	2.00	3.00	4.00	12.00	2.00

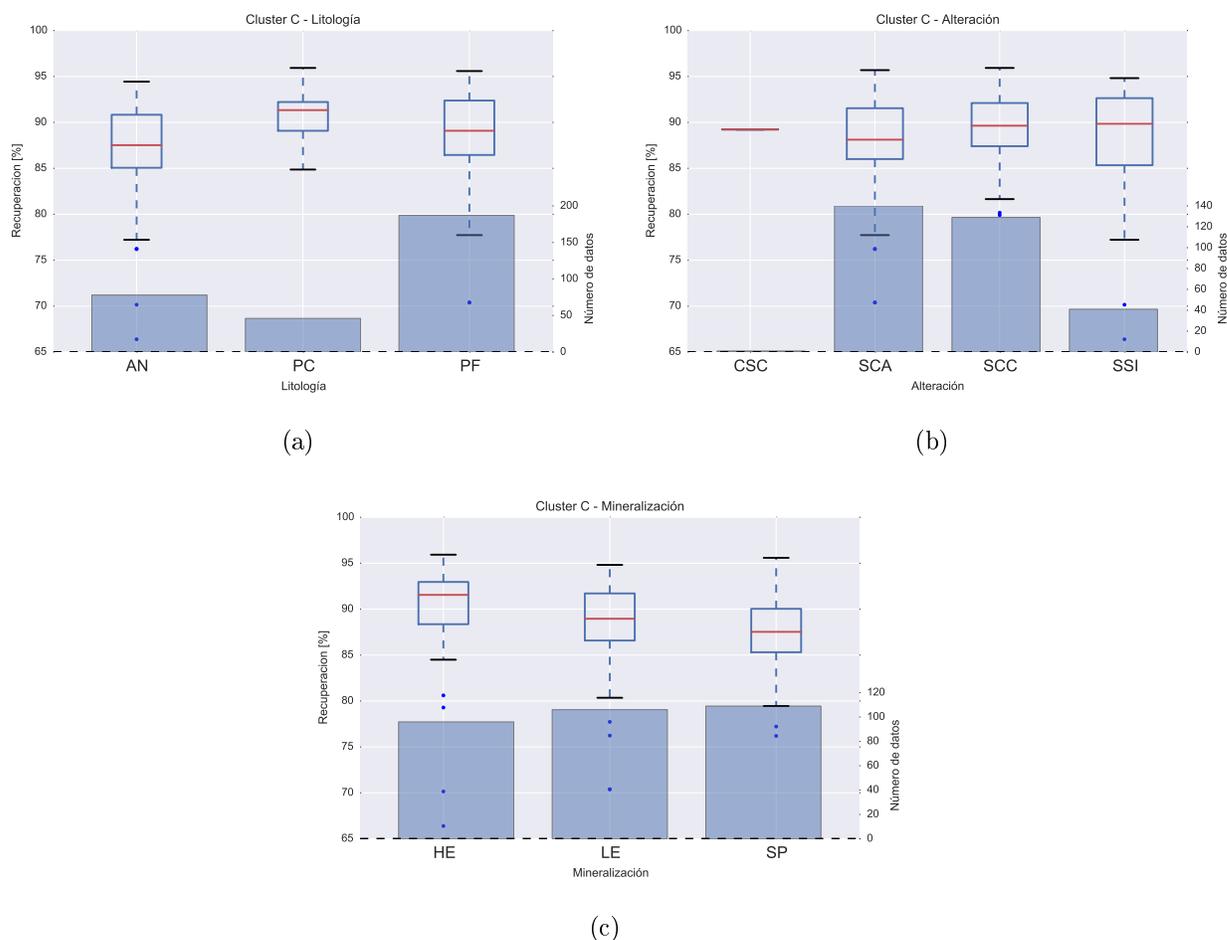


Figura 5.11: Boxplots de recuperación por tipo de variable geológica en el cluster C definido por k-Means. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.

5.3.1.4. Estadísticas cluster D de k-Means

Este cluster contiene 406 datos y sus estadísticas básicas se detallan en la Tabla 5.7.

Tabla 5.7: Estadísticas cluster D de k-Means

	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
$(Al+K)/(Na+Ca+Mg)$	6.13	2.86	0.47	0.65	3.92	6.27	7.75	16.33	3.83
$3*Al/(Na+K)$	6.27	1.69	0.27	3.57	5.13	5.85	6.95	13.88	1.82
Al	0.51	0.19	0.37	0.10	0.37	0.48	0.60	1.29	0.23
$Al/(Na+Ca+K)$	2.53	1.49	0.59	0.17	1.50	2.33	3.20	10.50	1.70
Al/Mg	17.45	6.33	0.36	2.50	13.33	16.50	20.33	48.00	7.00
Be	0.10	0.05	0.48	0.05	0.07	0.09	0.12	0.35	0.05
Cu	0.70	0.23	0.33	0.20	0.52	0.70	0.96	1.00	0.44
K/Al	0.42	0.11	0.25	0.15	0.35	0.43	0.50	0.69	0.15
K/Mg	7.18	2.80	0.39	0.59	5.30	6.71	8.50	21.00	3.20
Mg	0.03	0.02	0.68	0.01	0.02	0.03	0.04	0.31	0.02
P	113.77	44.73	0.39	20.00	80.00	110.00	130.00	320.00	50.00
S	2.38	1.43	0.60	0.67	1.55	2.08	2.75	10.00	1.20
V	3.47	2.42	0.70	1.00	2.00	3.00	4.00	19.00	2.00

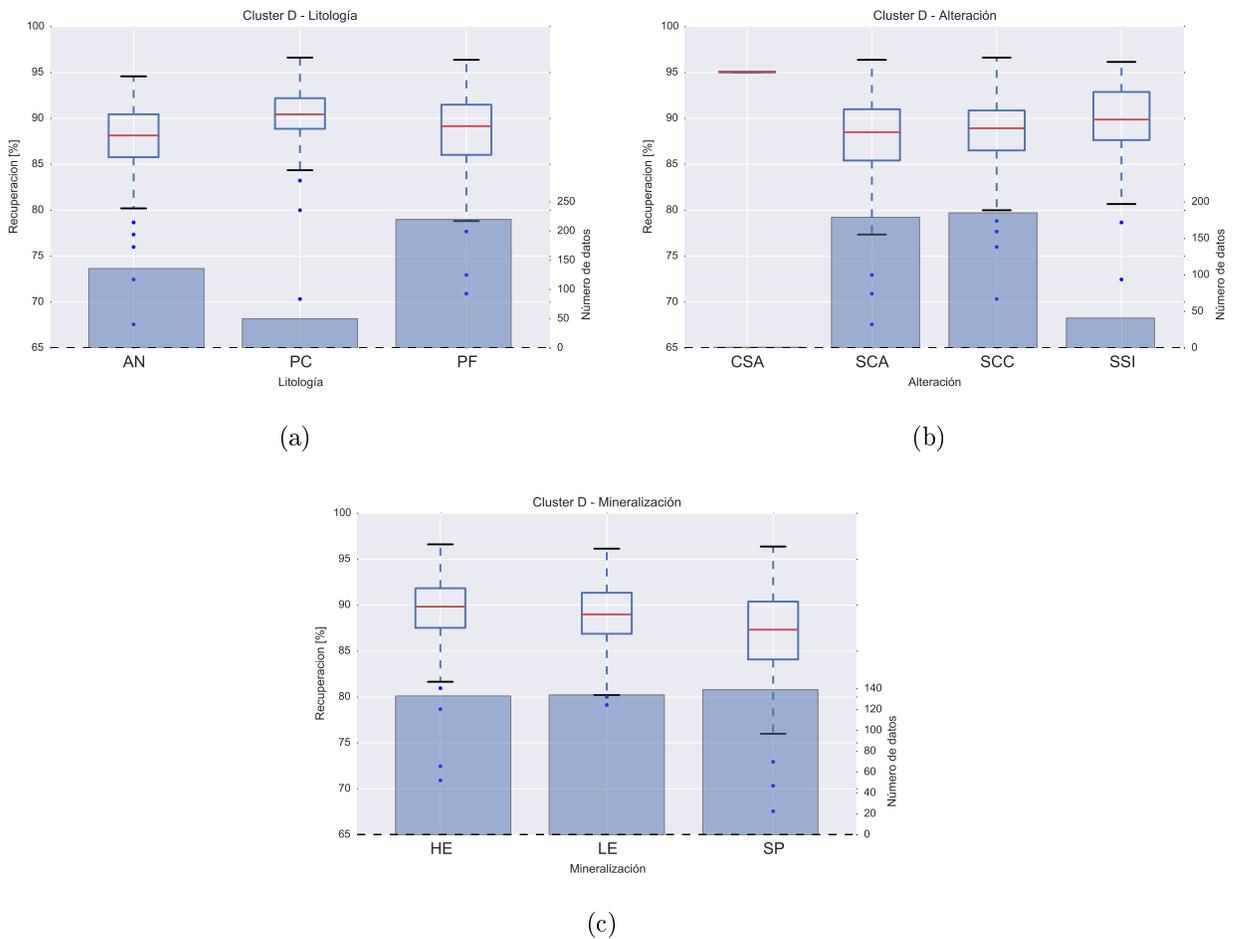


Figura 5.12: Boxplots de recuperación por tipo de variable geológica en el cluster D definido por k-Means. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.

5.3.2. Ward

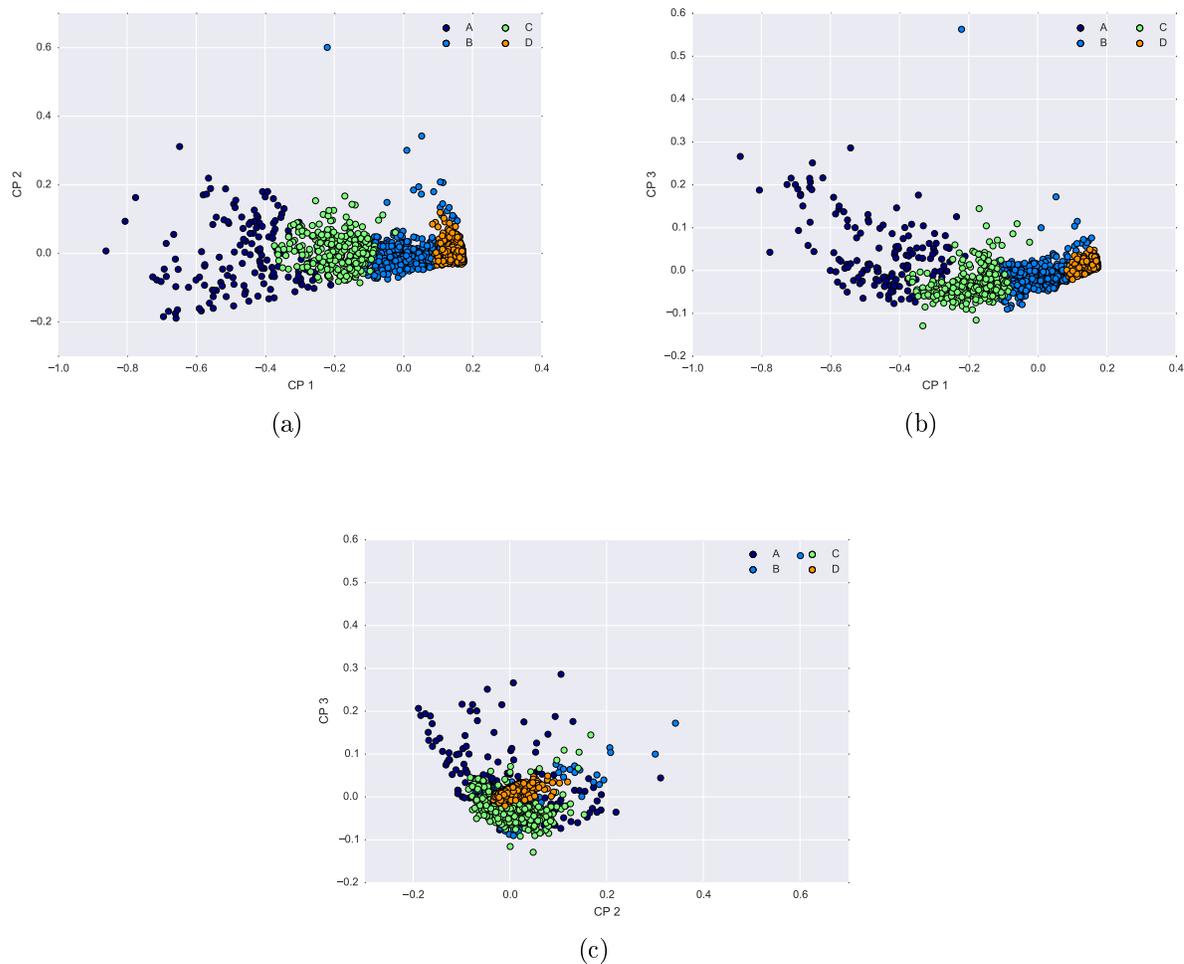


Figura 5.13: Resultados de clustering (Ward de 4 dominios) sobre datos de PCA. (a): Vista bidimensional componentes 1 y 2. (b): Vista bidimensional componentes 1 y 3. (c): Vista bidimensional componentes 2 y 3.

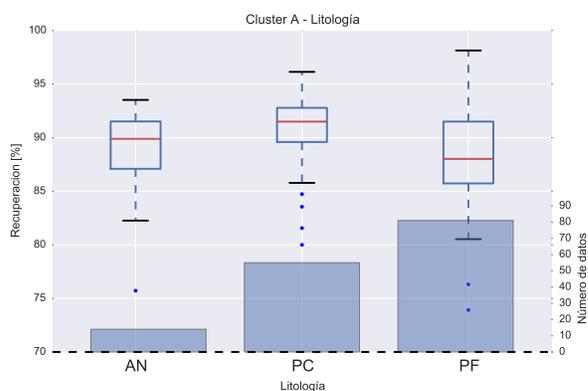
A continuación se mostrarán estadísticas básicas de cada uno de los clusters definidos a través de Ward.

5.3.2.1. Cluster A

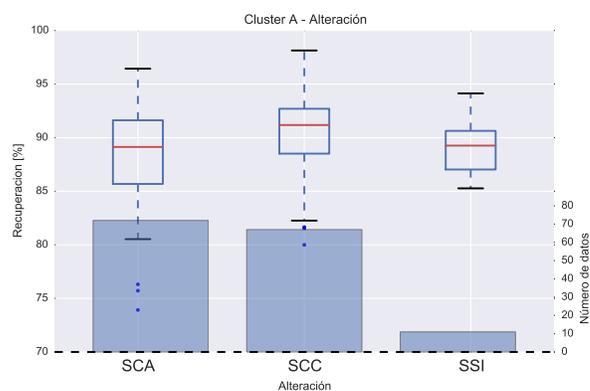
Este cluster contiene 150 datos y sus estadísticas básicas se detallan en la Tabla 5.8.

Tabla 5.8: Estadísticas cluster A de Ward

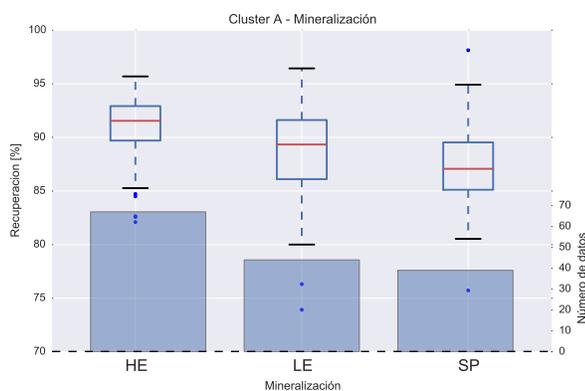
	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
$(Al+K)/(Na+Ca+Mg)$	5.22	3.33	0.64	0.29	2.30	4.87	7.35	16.00	5.06
$3*Al/(Na+K)$	6.00	2.19	0.36	3.75	4.45	5.16	6.90	16.80	2.46
Al	0.41	0.20	0.48	0.10	0.26	0.37	0.52	1.19	0.26
$Al/(Na+Ca+K)$	1.97	1.36	0.69	0.05	0.87	1.65	2.79	6.00	1.92
Al/Mg	21.36	7.24	0.34	9.50	16.13	20.50	24.25	48.00	8.13
Be	0.07	0.02	0.35	0.05	0.05	0.06	0.08	0.16	0.03
Cu	0.74	0.26	0.35	0.23	0.48	0.81	1.00	1.00	0.52
K/Al	0.46	0.14	0.31	0.14	0.35	0.49	0.58	0.70	0.23
K/Mg	9.81	4.85	0.49	2.00	5.54	9.00	12.00	27.00	6.46
Mg	0.02	0.01	0.47	0.01	0.01	0.02	0.03	0.04	0.02
P	34.20	13.72	0.40	10.00	30.00	30.00	40.00	90.00	10.00
S	2.71	1.84	0.68	0.88	1.65	2.13	3.09	10.00	1.44
V	2.81	1.63	0.58	1.00	2.00	2.00	4.00	9.00	2.00



(a)



(b)



(c)

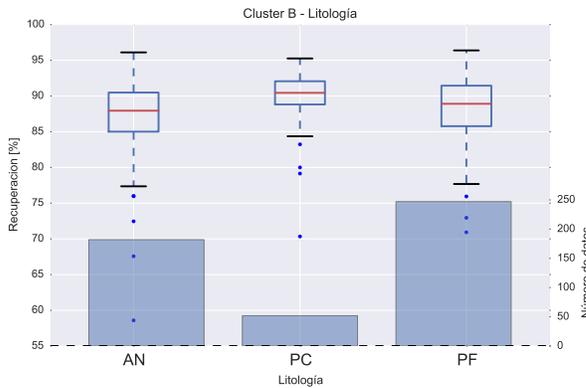
Figura 5.14: Boxplots de recuperación por tipo de variable geológica en el cluster A definido por Ward. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.

5.3.2.2. Cluster B

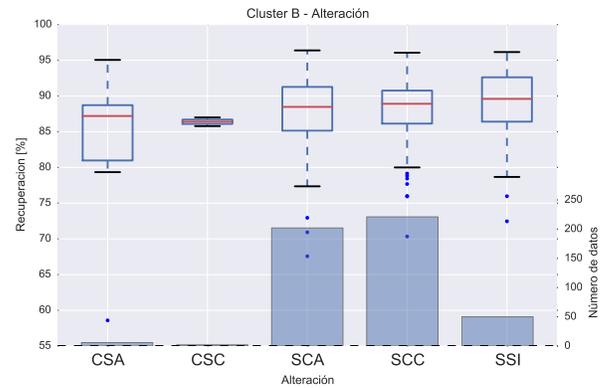
Este cluster contiene 481 datos y sus estadísticas básicas se detallan en la Tabla 5.9.

Tabla 5.9: Estadísticas cluster B de Ward

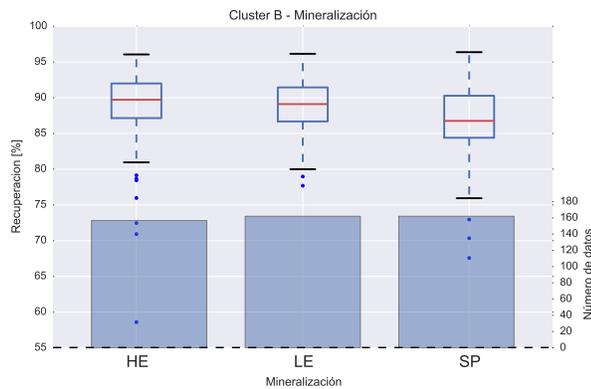
	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
$(Al+K)/(Na+Ca+Mg)$	5.85	2.79	0.48	0.52	3.65	6.00	7.50	16.33	3.85
$3*Al/(Na+K)$	6.50	2.09	0.32	3.57	5.22	5.88	7.13	22.56	1.91
Al	0.54	0.26	0.49	0.15	0.38	0.48	0.61	3.02	0.23
$Al/(Na+Ca+K)$	2.43	1.43	0.59	0.13	1.40	2.29	3.10	10.50	1.70
Al/Mg	16.76	6.51	0.39	1.14	13.00	16.50	20.00	48.00	7.00
Be	0.11	0.06	0.56	0.05	0.07	0.09	0.13	0.49	0.06
Cu	0.71	0.23	0.33	0.20	0.52	0.71	0.97	1.00	0.45
K/Al	0.41	0.11	0.27	0.09	0.33	0.42	0.49	0.69	0.16
K/Mg	6.81	2.92	0.43	0.16	5.00	6.50	8.50	21.00	3.50
Mg	0.05	0.15	2.74	0.01	0.02	0.03	0.04	1.82	0.02
P	131.62	69.63	0.53	20.00	90.00	110.00	160.00	900.00	70.00
S	2.41	1.48	0.61	0.67	1.55	2.08	2.81	10.00	1.26
V	4.98	9.53	1.91	1.00	2.00	3.00	5.00	168.00	3.00



(a)



(b)



(c)

Figura 5.15: Boxplots de recuperación por tipo de variable geológica en el cluster B definido por Ward. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.

5.3.2.3. Cluster C

Este cluster contiene 301 datos y sus estadísticas básicas se detallan en la Tabla 5.10.

Tabla 5.10: Estadísticas cluster C de Ward

	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
$(Al+K)/(Na+Ca+Mg)$	7.09	2.98	0.42	0.50	5.44	7.11	9.00	14.75	3.56
$3*Al/(Na+K)$	6.58	2.20	0.33	3.63	5.05	5.81	7.78	19.32	2.74
Al	0.51	0.21	0.42	0.08	0.35	0.46	0.64	1.61	0.29
$Al/(Na+Ca+K)$	2.77	1.36	0.49	0.05	2.00	2.71	3.57	10.00	1.57
Al/Mg	19.19	6.41	0.33	4.83	15.00	18.00	22.67	47.00	7.67
Be	0.09	0.04	0.44	0.05	0.05	0.08	0.11	0.29	0.06
Cu	0.75	0.23	0.31	0.29	0.56	0.78	1.00	1.00	0.44
K/Al	0.41	0.12	0.31	0.11	0.31	0.42	0.50	0.70	0.19
K/Mg	7.53	2.93	0.39	1.00	5.50	7.00	9.00	19.00	3.50
Mg	0.03	0.01	0.34	0.01	0.02	0.03	0.03	0.06	0.01
P	60.90	21.03	0.35	30.00	50.00	60.00	70.00	150.00	20.00
S	2.29	1.48	0.65	0.73	1.47	1.88	2.60	10.00	1.13
V	2.99	1.70	0.57	1.00	2.00	3.00	4.00	12.00	2.00

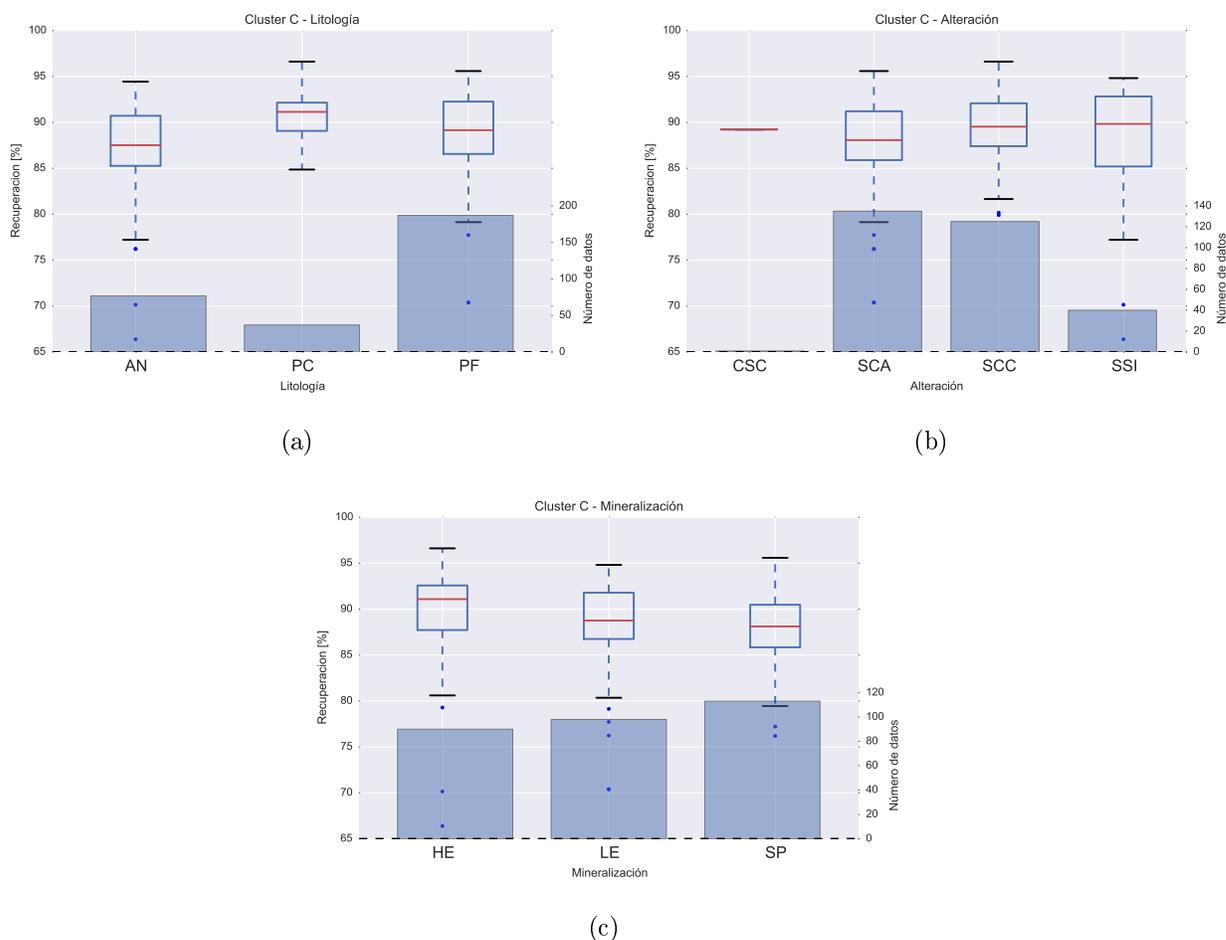


Figura 5.16: Boxplots de recuperación por tipo de variable geológica en el cluster C definido por Ward. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.

5.3.2.4. Cluster D

Este cluster contiene 774 datos y sus estadísticas básicas se detallan en la Tabla 5.11.

Tabla 5.11: Estadísticas cluster D de Ward

	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
$(Al+K)/(Na+Ca+Mg)$	2.23	1.61	0.72	0.13	1.14	1.53	2.83	11.38	1.69
$3*Al/(Na+K)$	11.57	5.47	0.47	1.85	7.41	9.91	14.28	35.67	6.88
Al	1.20	0.65	0.55	0.16	0.64	1.00	1.70	3.81	1.06
$Al/(Na+Ca+K)$	1.20	0.72	0.60	0.05	0.64	1.00	1.61	5.00	0.97
Al/Mg	4.07	4.91	1.21	0.24	1.33	1.78	4.03	30.25	2.70
Be	0.42	0.23	0.55	0.05	0.26	0.38	0.55	1.56	0.29
Cu	0.66	0.22	0.33	0.17	0.50	0.64	0.85	1.00	0.35
K/Al	0.24	0.10	0.43	0.05	0.16	0.23	0.31	0.63	0.15
K/Mg	1.20	1.72	1.43	0.09	0.22	0.38	1.22	10.33	1.01
Mg	0.72	0.60	0.84	0.02	0.16	0.53	1.25	2.39	1.09
P	729.77	466.57	0.64	70.00	302.50	630.00	1140.00	2910.00	837.50
S	1.64	1.00	0.61	0.37	1.02	1.37	2.00	7.93	0.98
V	25.53	22.08	0.86	1.00	7.00	17.00	40.00	118.00	33.00

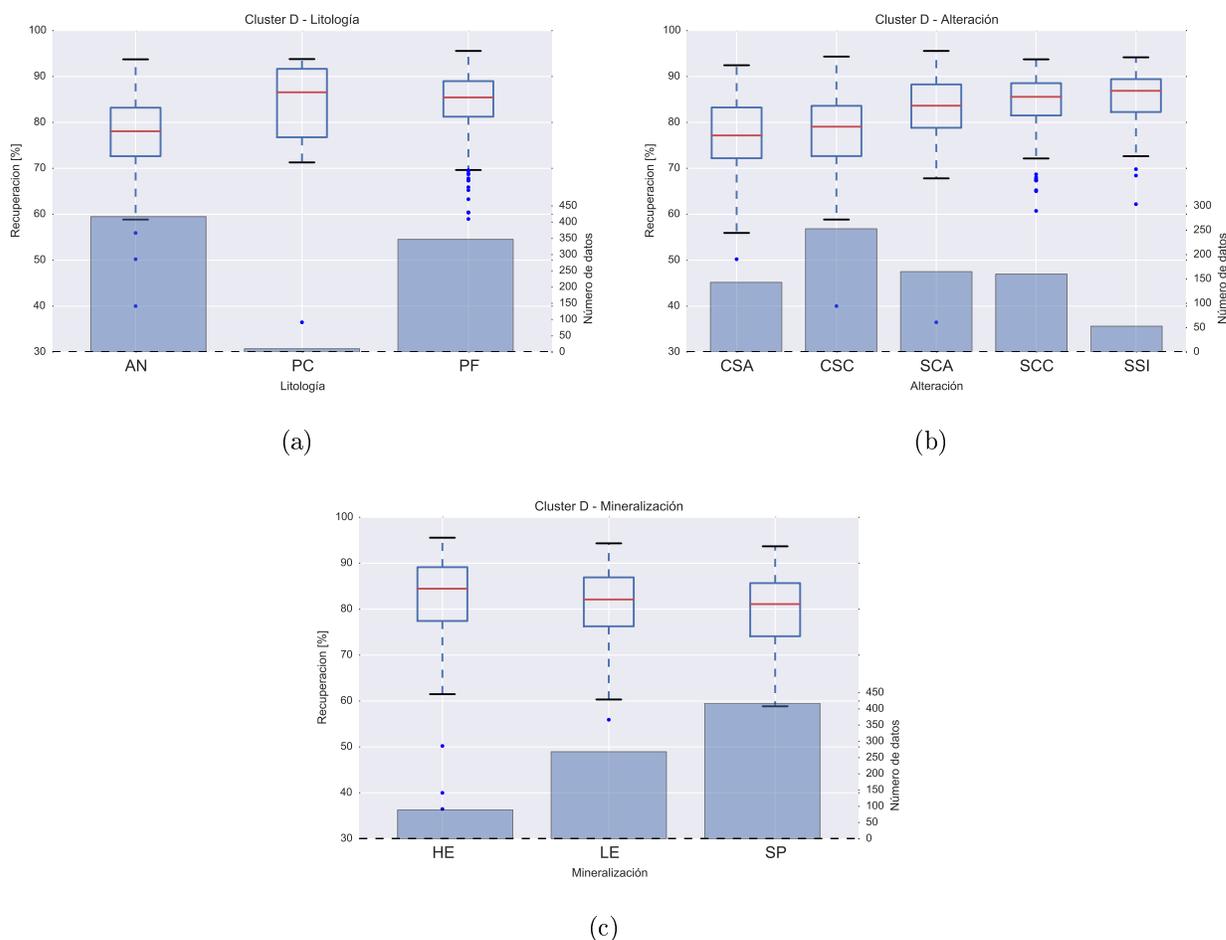


Figura 5.17: Boxplots de recuperación por tipo de variable geológica en el cluster D definido por Ward. (a): Recuperación por litologías. (b): Recuperación por alteraciones. (c): Recuperación por mineralizaciones.

5.3.3. Discusiones

5.3.3.1. Semejanza entre clusters definidos

En primer lugar se compararán los clusters generados por k-Means con respecto a los de Ward. La herramienta idónea para esto es una matriz de confusión, la cual compara la cantidad de datos (porcentaje) existentes en la clasificación realizada por un método (k-Means) que coinciden con la clasificación realizada por otro (Ward).

Tabla 5.12: Matriz de confusión entre resultados de k-Means y Ward.

		Ward			
		A	B	C	D
k-Means	A	0 %	10 %	0 %	90 %
	B	96 %	1 %	3 %	0 %
	C	10 %	1 %	89 %	0 %
	D	0 %	95 %	5 %	0 %

A partir de los resultados en la Tabla 5.12, se obtienen las siguientes relaciones entre los dominios obtenidos:

- Cluster A de k-Means es semejante en un 90 % con el cluster D de Ward.
- Cluster B de k-Means es semejante en un 96 % con el cluster A de Ward.
- Cluster C de k-Means es semejante en un 89 % con el cluster C de Ward.
- Cluster D de k-Means es semejante en un 95 % con el cluster B de Ward.

Se observa que los grupos definidos por k-Means presentan una gran semejanza con los obtenidos a través de Ward, por lo que se esperan resultados semejantes en cuanto a sus estadísticas tanto en las variables geoquímicas/sintéticas como en recuperaciones.

5.3.3.2. Recuperación

En cuanto a las recuperaciones (Figura 5.18), las distribuciones de recuperación por cada cluster definido por k-Means son prácticamente las mismas que sus símiles en Ward, por lo que la decisión sobre cual método elegir dependerá del criterio del evaluador. Si se considerase como criterio el coeficiente de silueta (Figura 5.7), el método a elegir sería k-Means ya que posee un mayor coeficiente que Ward, lo que indica que agrupa mejor los datos.

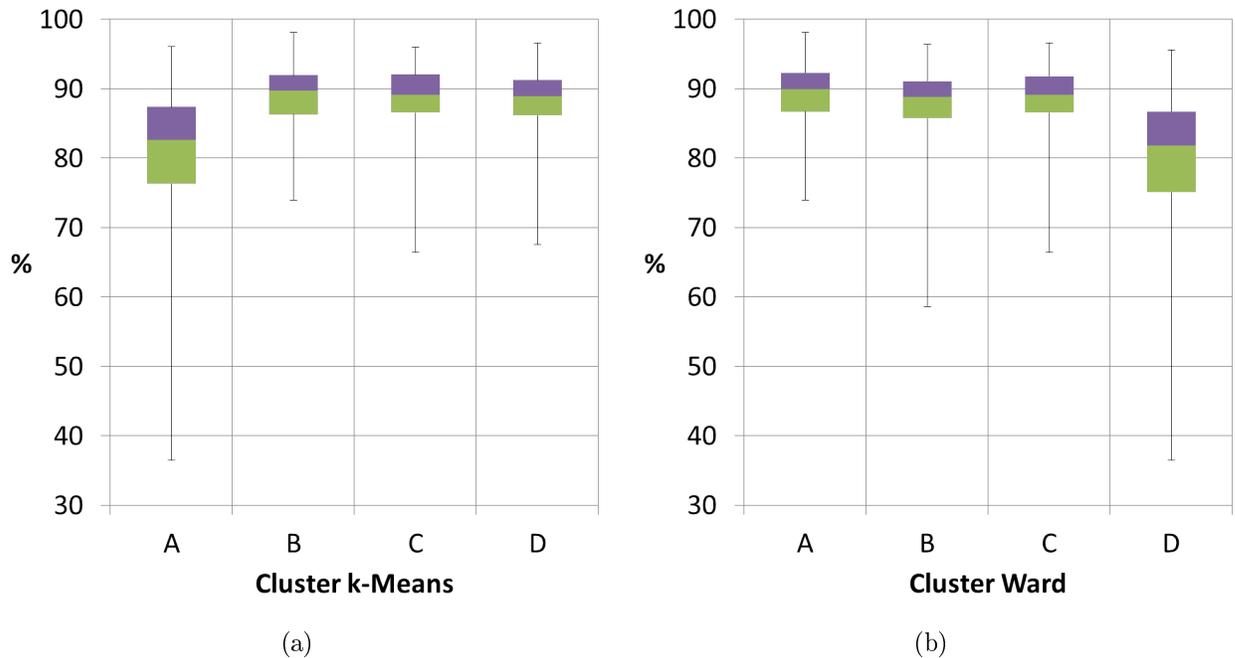


Figura 5.18: Boxplots de recuperación por cluster en cada método. (a): k-Means. (b): Ward.

5.3.3.3. Separación geológica

Una vez definidos los clusters, es importante estudiar e identificar las asociaciones geológicas presentes en ellas, tanto en litologías, alteraciones y mineralizaciones. La finalidad de este análisis reside en la comprensión de la calidad de clasificación y si realmente existe una real separación que permita identificar diferencias significativas entre un dominio y otro.

Para aliviar el análisis de los clusters definidos, se realizará el estudio sobre los resultados obtenidos a través de k-Means únicamente debido a la alta semejanza con respecto a los dominios obtenidos a través de Ward (Sección 5.3.3.1).

Litologías

Se observa que el cluster A está asociado a un alto porcentaje de Andesita (AN) y uno muy bajo de Pórfido Cuarzófero (PC). El cluster B, al contrario del A, presenta una alta proporción de PC y la más baja de AN. En cuanto a los clusters C y D, estos no presentan mayor diferencia salvo disimilitudes en la asociación de Andesita, por lo que dependiendo del criterio, estos clusters pueden ser considerados redundantes.

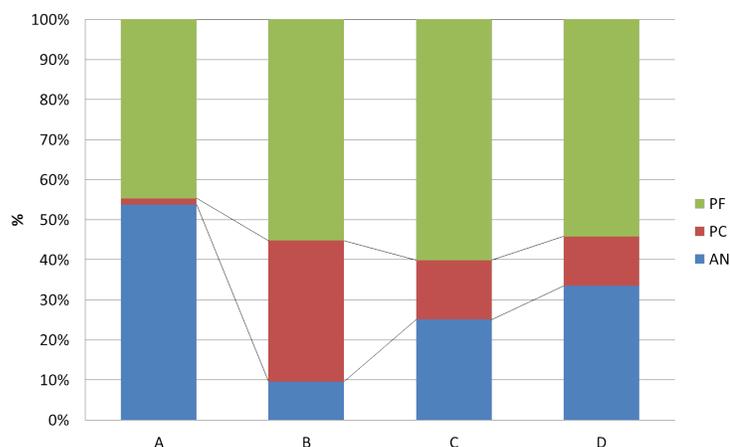


Figura 5.19: Proporciones de litologías para cada cluster definido por k-Means.

Alteraciones

En cuanto a las alteraciones, en todos los clusters existe una equidistribución de las Sericita-Cuarzo con y sin arcillas (SCC y SCA). Sin embargo, el cluster A agrupa prácticamente la totalidad de las alteraciones Clorita-Sericita-Cuarzo (CSC) y Clorita-Sericita-Arcillas (CSA), indicando que este dominio es el que contiene mayor cantidad de arcillas. Se puede concluir que los clusters B, C y D presentan proporciones muy semejantes entre sí, por lo que pueden ser considerados redundantes.

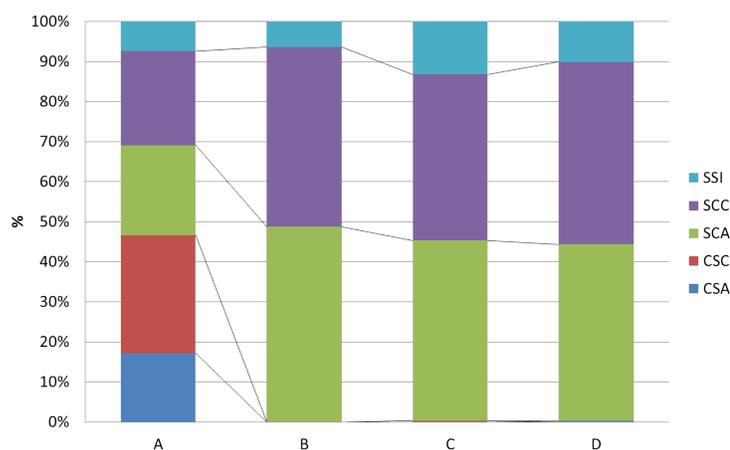


Figura 5.20: Proporciones de alteraciones para cada cluster definido por k-Means.

Mineralizaciones

Finalmente, con respecto a las mineralizaciones, se observa que el cluster A se diferencia por poseer una alta proporción de Sulfuros Primarios (SP) y una baja de minerales de alto enriquecimiento (HE). El cluster B se caracteriza por un alto porcentaje de HE y uno relativamente bajo de SP. Los dos cluster restantes (C y D) presentan proporciones muy

semejantes, por lo que deben ser considerados redundantes al clasificar el mismo tipo de información. Finalmente, se destaca que ningún dominio definido diferencia en mayor medida los minerales de bajo enriquecimiento (LE).

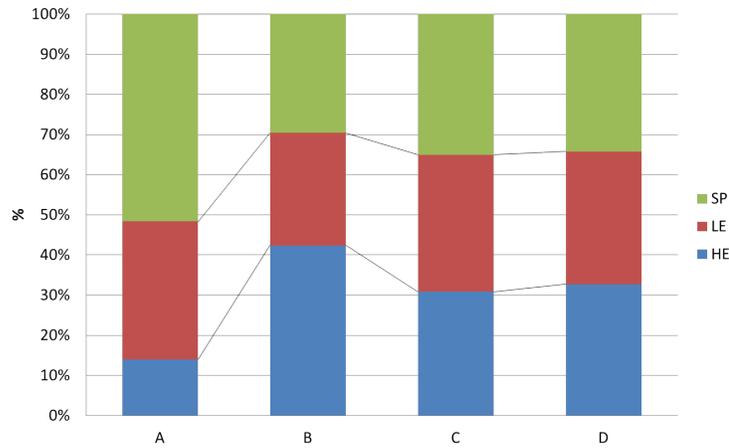


Figura 5.21: Proporciones de mineralizaciones para cada cluster definido por k-Means.

5.4. Modelo predictivo

Una vez definidos los dominios, se realizó un modelo predictivo de la recuperación de cobre en cada uno de ellos a través de regresión multilínea. Tal como se mencionó en los antecedentes, el método utilizado para obtener las regresiones es LARS Lasso, el que dependiendo del valor de un coeficiente s permite seleccionar variables al realizar el modelo predictivo, con una consecuente pérdida de precisión en la predicción (menor coeficiente de determinación).

Se consideraron 5 valores para este coeficiente s : 0; 0.01; 0.025; 0.05; 0.075 y 0.1. En donde el valor 0 entrega resultados equivalentes a realizar un simple modelo de regresión multilínea a través de mínimos cuadrados y, a medida que aumenta s , se omiten variables con menor ponderación en la regresión, disminuyendo la cantidad de variables con el costo de reducir también su valor de R^2 asociado. Finalmente, se aplica LARS Lasso en cada uno de los 4 dominios obtenidos con k-Means y Ward, para cada valor de s .

Es importante destacar que el objetivo del modelo predictivo es encontrar una relación multilínea que prediga de la mejor forma posible la recuperación de cobre, con la menor cantidad de variables posible, esto es para evitar sobreajuste y problemas al aplicar este modelo sobre datos nuevos.

5.4.1. k-Means

5.4.1.1. Cluster A

Se observa en la Tabla 5.13 que a un valor de $s = 0.025$ se obtiene un modelo predictivo que considera solo 5 variables, con un R^2 de 0.492 correspondiente a una correlación de Pearson de 0.701.

Tabla 5.13: Modelo predictivo en Cluster A de k-Means

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(Al+K)/(Na+Ca+Mg)	-0.817	0	0	0	0	0
3*Al/(Na+K)	-0.395	-0.355	-0.333	-0.278	-0.200	-0.122
Al	-4.267	-2.443	-2.589	-2.870	-2.766	-2.661
Al/(Na+Ca+K)	1.952	0.582	0.275	0	0	0
Al/Mg	0.147	0	0	0	0	0
Be	-7.720	-7.381	-6.376	-4.585	-2.731	-0.877
Cu	3.851	3.372	1.913	0	0	0
K/Al	-8.466	0	0	0	0	0
K/Mg	0.215	0.037	0	0	0	0
Mg	0.087	0	0	0	0	0
P	-0.0002	0	0	0	0	0
S	0.280	0.026	0	0	0	0
V	0.039	0	0	0	0	0
Intercepto	90.189	88.026	89.036	89.456	87.738	86.020
R^2	0.518	0.506	0.492	0.449	0.397	0.325
N° de datos	864	864	864	864	864	864

5.4.1.2. Cluster B

En este dominio, con un valor de $s = 0.025$ el modelo predictivo consta de 7 variables y un R^2 de 0.392, en cambio con $s = 0.05$ el modelo se reduce a 4 variables y un R^2 de 0.327.

Tabla 5.14: Modelo predictivo en Cluster B de k-Means

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(Al+K)/(Na+Ca+Mg)	-1.510	0	0	0	0	0
3*Al/(Na+K)	-0.136	-0.228	-0.389	-0.334	-0.237	-0.126
Al	-9.823	-7.197	-7.427	-6.188	-5.002	-3.971
Al/(Na+Ca+K)	3.349	0.354	0.122	0	0	0
Al/Mg	0.886	0	0	0	0	0
Be	9.009	1.278	0	0	0	0
Cu	6.437	5.960	4.866	3.025	1.119	0
K/Al	24.758	2.694	0	0	0	0
K/Mg	-1.451	0	0	0	0	0
Mg	58.153	-20.465	0	0	0	0
P	-0.059	-0.035	-0.003	0	0	0
S	0.252	0.205	0.072	0	0	0
V	-0.241	-0.256	-0.179	-0.063	0	0
Intercepto	74.492	88.587	90.983	91.511	91.678	91.415
R^2	0.492	0.426	0.392	0.327	0.246	0.173
N° de datos	125	125	125	125	125	125

5.4.1.3. Cluster C

Existe un pobre coeficiente de determinación en general para este cluster. Con un valor de $s = 0.01$ el modelo queda reducido a 6 variables con un $R^2 = 0.261$. Aumentando el valor de s se generan modelos con valores muy bajos de R^2 , por lo que no se aconseja utilizar modelos para este cluster con valores de s superiores a 0.01.

Tabla 5.15: Modelo predictivo en Cluster C de k-Means

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(Al+K)/(Na+Ca+Mg)	-0.310	0	0	0	0	0
3*Al/(Na+K)	0.632	0	0	0	0	0
Al	1.740	0	0	0	0	0
Al/(Na+Ca+K)	0.576	0	0	0	0	0
Al/Mg	0.023	0	0	0	0	0
Be	16.571	3.934	0	0	0	0
Cu	9.127	7.969	6.308	3.776	1.519	0
K/Al	10.020	2.664	1.308	0	0	0
K/Mg	0.012	0	0	0	0	0
Mg	-150.651	-112.994	-92.924	-39.221	0	0
P	-0.015	0	0	0	0	0
S	0.386	0.091	0	0	0	0
V	-0.446	-0.126	-0.001	0	0	0
Intercepto	76.765	84.510	85.948	86.942	87.575	88.698
R^2	0.288	0.261	0.226	0.145	0.052	0
N° de datos	311	311	311	311	311	311

5.4.1.4. Cluster D

En este dominio ocurre un caso muy semejante al Cluster C, donde a un bajo valor de s ya se tiene un bajo coeficiente de determinación. En este caso, con $s = 0.01$ se obtiene un modelo que relaciona 5 variables a un R^2 de solo 0.167.

Tabla 5.16: Modelo predictivo en Cluster D de k-Means

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(Al+K)/(Na+Ca+Mg)	-0.372	0	0	0	0	0
3*Al/(Na+K)	1.011	0	0	0	0	0
Al	-5.167	-2.087	-1.535	0	0	0
Al/(Na+Ca+K)	0.491	0	0	0	0	0
Al/Mg	0.119	0	0	0	0	0
Be	8.679	0	0	0	0	0
Cu	6.290	5.222	3.739	1.509	0	0
K/Al	11.148	0	0	0	0	0
K/Mg	0.055	0.168	0.064	0	0	0
Mg	-2.091	0	0	0	0	0
P	-0.002	0	0	0	0	0
S	0.243	0.071	0	0	0	0
V	-0.199	-0.062	0	0	0	0
Intercepto	73.598	84.598	86.058	87.302	88.360	88.360
R^2	0.203	0.167	0.127	0.047	0	0
N° de datos	406	406	406	406	406	406

5.4.2. Ward

5.4.2.1. Cluster A

Se observa en la Tabla 5.17 que a un valor de $s = 0.025$ el modelo queda reducido a 5 variables con un R^2 de 0.312. Considerando $s = 0.01$ mejora el valor de R^2 a 0.355, sin embargo, el número de variables aumenta a 9.

Tabla 5.17: Modelo predictivo en Cluster A de Ward

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(Al+K)/(Na+Ca+Mg)	-0.973	0	0	0	0	0
3*Al/(Na+K)	0.072	0	0	-0.023	0	0
Al	-12.784	-8.609	-7.146	-6.547	-4.672	-2.639
Al/(Na+Ca+K)	2.241	0.181	0	0	0	0
Al/Mg	0.806	0	0	0	0	0
Be	27.116	17.249	0	0	0	0
Cu	6.680	6.445	5.156	3.006	1.273	0
K/Al	26.390	2.525	1.772	0	0	0
K/Mg	-1.391	-0.011	0	0	0	0
Mg	45.892	-27.099	-5.366	0	0	0
P	-0.026	0	0	0	0	0
S	0.161	0.070	0	0	0	0
V	-0.269	-0.254	-0.204	-0.024	0	0
Intercepto	72.634	86.576	88.293	89.973	90.275	90.380
R^2	0.424	0.355	0.312	0.250	0.173	0.087
N° de datos	150	150	150	150	150	150

5.4.2.2. Cluster B

Se observa que un modelo multilineal ajusta pobremente los datos de este cluster. A un valor de $s = 0.01$ se obtienen 6 variables con un coeficiente de determinación de solo 0.192.

Tabla 5.18: Modelo predictivo en Cluster B de Ward

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(Al+K)/(Na+Ca+Mg)	0.081	0	0	0	0	0
3*Al/(Na+K)	0.399	0	0	0	0	0
Al	-1.204	0	-0.347	-0.033	0	0
Al/(Na+Ca+K)	-0.206	0	0	0	0	0
Al/Mg	0.117	0	0	0	0	0
Be	0.052	-0.404	-0.344	0	0	0
Cu	6.051	5.120	3.590	0.983	0	0
K/Al	13.824	2.921	1.766	0	0	0
K/Mg	-0.057	0.180	0.102	0	0	0
Mg	2.714	0	0	0	0	0
P	0.001	0	0	0	0	0
S	0.296	0.164	0	0	0	0
V	-0.100	-0.056	-0.029	0	0	0
Intercepto	74.126	81.979	84.512	87.440	88.123	0
R^2	0.209	0.192	0.148	0.029	0.000	0
N° de datos	481	481	481	481	481	481

5.4.2.3. Cluster C

Se observa un bajo coeficiente de determinación en el ajuste del modelo multilineal. En el caso de $s = 0.025$, se tiene un modelo con 3 variables y un valor de R^2 igual a 0.206.

Tabla 5.19: Modelo predictivo en Cluster C de Ward

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(Al+K)/(Na+Ca+Mg)	-0.492	0	0	0	0	0
3*Al/(Na+K)	0.775	0	0	0	0	0
Al	3.798	0	0	0	0	0
Al/(Na+Ca+K)	0.802	0	0	0	0	0
Al/Mg	-0.073	0	0	0	0	0
Be	18.437	6.175	0	0	0	0
Cu	9.190	7.690	5.943	3.367	1.106	0
K/Al	8.998	3.269	1.817	0	0	0
K/Mg	0.170	0.030	0	0	0	0
Mg	-186.762	-111.331	-90.479	-38.307	0	0
P	-0.002	0	0	0	0	0
S	0.453	0.168	0	0	0	0
V	-0.447	-0.083	0	0	0	0
Intercepto	76.364	83.648	85.869	87.102	87.740	88.564
R^2	0.275	0.245	0.206	0.124	0.035	0.000
N° de datos	301	301	301	301	301	301

5.4.2.4. Cluster D

Se observa en la Tabla 5.20 que a un valor de $s = 0.025$ se obtiene un modelo predictivo que considera solo 5 variables, con un R^2 de 0.490.

Tabla 5.20: Modelo predictivo en Cluster D de Ward

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(Al+K)/(Na+Ca+Mg)	-1.183	0	0	0	0	0
3*Al/(Na+K)	-0.351	-0.339	-0.316	-0.262	-0.188	-0.114
Al	-5.537	-3.009	-3.167	-3.289	-3.190	-3.090
Al/(Na+Ca+K)	2.333	0.660	0.296	0	0	0
Al/Mg	0.189	0	0	0	0	0
Be	-7.256	-6.944	-5.726	-3.734	-1.902	-0.071
Cu	3.244	3.051	1.592	0	0	0
K/Al	-9.824	0	0	0	0	0
K/Mg	0.421	0	0	0	0	0
Mg	0.198	0	0	0	0	0
P	0.000	0	0	0	0	0
S	0.318	0	0	0	0	0
V	0.054	0	0	0	0	0
Intercepto	91.042	88.220	89.035	89.132	87.377	85.622
R^2	0.505	0.490	0.477	0.439	0.392	0.325
N° de datos	774	774	774	774	774	774

5.4.3. Discusiones

Previo a la discusión de los resultados obtenidos en los modelos predictivos, es necesario recordar la semejanza entre los cuatro clusters obtenidos por cada método. A partir de los resultados de la matriz de confusión (Tabla 5.12) se determinaron las siguientes similitudes:

- Cluster A de k-Means es semejante en un 90 % con el cluster D de Ward.
- Cluster B de k-Means es semejante en un 96 % con el cluster A de Ward.
- Cluster C de k-Means es semejante en un 89 % con el cluster C de Ward.
- Cluster D de k-Means es semejante en un 95 % con el cluster B de Ward.

Al ser altos porcentajes de semejanza, se puede establecer paralelismo entre los resultados de regresión obtenidos para cada modelo. En base a los resultados obtenidos se llegó a las siguientes conclusiones:

- En los dominios con mayor coeficiente de determinación (A y B en k-Means; D y A en Ward), los valores más altos de R^2 se obtuvieron en los clusters definidos a través de k-Means.
- Al contrario, de los clusters que presentaban menor coeficiente de determinación (C y D en k-Means; B y C en Ward), los valores más altos fueron los obtenidos por Ward.

Con estos resultados a la vista, se puede afirmar que es responsabilidad del investigador evaluar los modelos predictivos obtenidos por cada método, así como identificar la cantidad óptima de variables a utilizar, con la menor pérdida de precisión posible.

5.5. Análisis de sensibilidad

El objetivo primordial de esta sección es comparar las proporciones de litologías, alteraciones y mineralizaciones en cada dominio, para los algoritmos de k-Means y Ward. La idea básica es concluir si añadir más clusters ayuda o no a separar estas variables geológicas, con la finalidad de identificar un número óptimo de clusters finales a utilizar, bajo la mirada crítica del analista que aplicará esta metodología.

En este caso se compararán los resultados, tanto de k-Means como de Ward, al definir 4, 5 y 6 clusters y se estudiarán las proporciones obtenidas de las litologías, alteraciones y mineralizaciones presentes.

5.5.1. 4 clusters

5.5.1.1. Clusters definidos

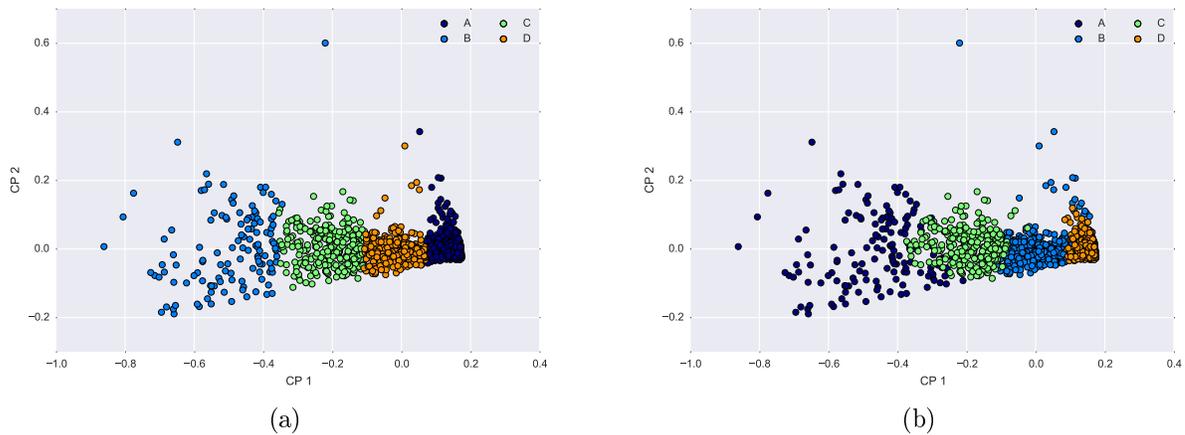


Figura 5.22: 4 clusters definidos por cada método (a) k-Means. (b) Ward

5.5.1.2. Litologías

Se observa en la Figura 5.23 que los resultados son semejantes entre los dos métodos. Existe un cluster (A en k-Means y D en Ward) que separa mayormente la Andesita (AN) y en muy baja cantidad Pórfido Cuprífero (PC), a diferencia de otro cluster (B en k-Means y A en Ward), que concentra una mayor cantidad de Pórfido Cuarzífero que la Andesita, sin embargo, no logra mayor diferenciación en el Pórfido Feldespático, en los dos casos.

Los dos clusters restantes no se diferencian en sus proporciones de Pórfido Cuarzífero. Sin embargo, presentan diferencias al clasificar Pórfido Feldespático y Andesita, sobre todo en los dominios obtenidos a través de Ward.

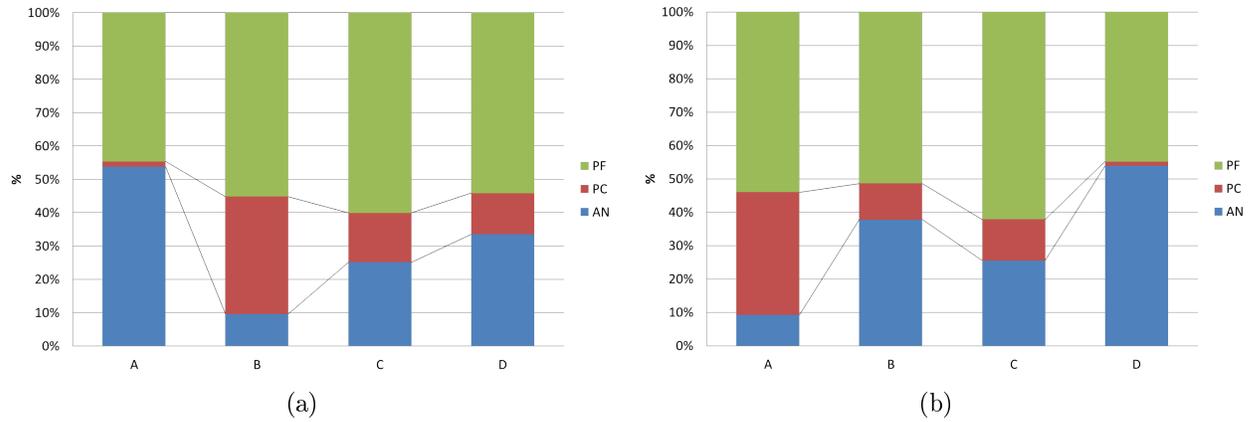


Figura 5.23: Proporciones de litologías por tipo de método para 4 clusters. (a) k-Means. (b) Ward

5.5.1.3. Alteraciones

Se observa en los dos métodos la generación de un cluster (A en k-Means y D en Ward) que agrupa la totalidad de la alteración Clorita-Sericita-Cuarzo (CSC) y la alta mayoría de la alteración Clorita-Sericita-Arcillas (CSA). El resto de los grupos son muy parecidos entre sí y agrupan las alteraciones Sericíticas, por lo que dentro de un determinado criterio pueden ser considerados redundantes y ser unidos dentro de un solo grupo o dominio.

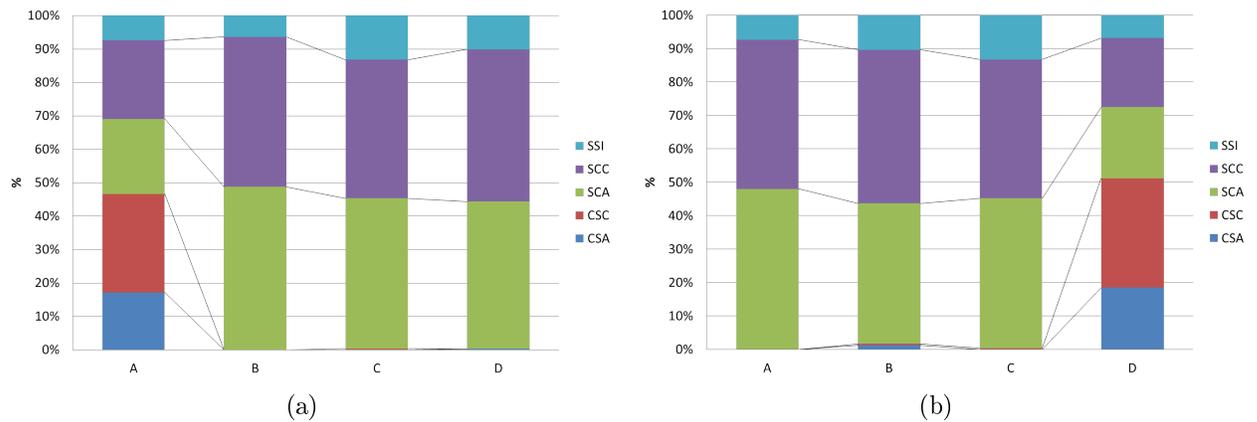


Figura 5.24: Proporciones de alteraciones por tipo de método para 4 clusters. (a) k-Means. (b) Ward

5.5.1.4. Mineralizaciones

En cuanto a las mineralizaciones, se obtiene nuevamente resultados muy semejantes entre los dos métodos, generando un par de clusters iguales entre sí (C y D en k-Means y B y C en Ward), con proporciones equitativas de los tres tipos de mineralización. Por otro lado, uno

de los grupos obtenidos (A en k-Means y D en Ward), presenta una mayor proporción de sulfuros primarios (SP) y agrupa pocos minerales de alto enriquecimiento (HE), junto con otro cluster (B en k-Means y A en Ward) que concentra mayor proporción de minerales asociados a la zona de alto enriquecimiento (HE). Finalmente, se observa que ningún grupo obtenido agrupa mayoritariamente minerales de bajo enriquecimiento, manteniendo proporciones relativamente constantes en los cuatro dominios.

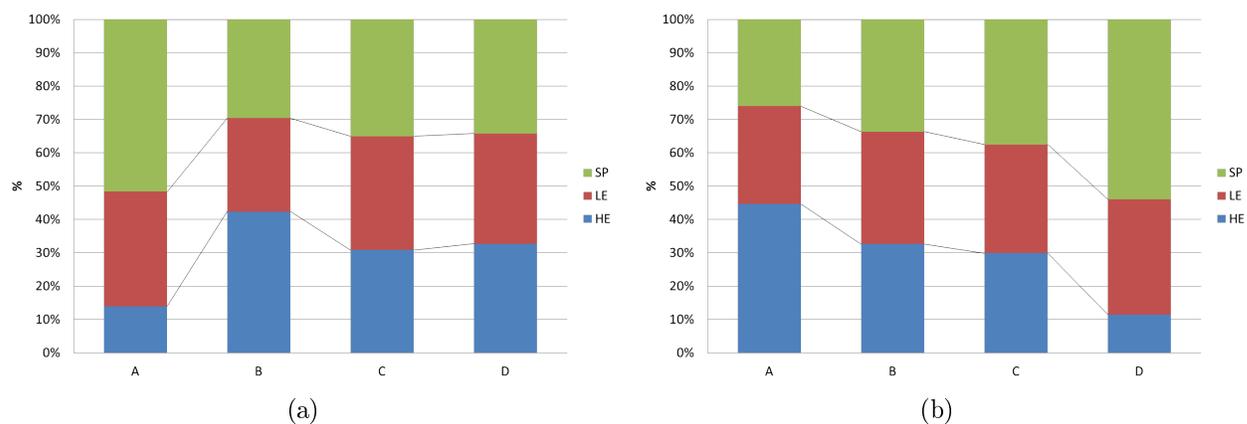


Figura 5.25: Proporciones de mineralizaciones por tipo de método para 4 clusters. (a) k-Means. (b) Ward

5.5.2. 5 clusters

5.5.2.1. Clusters definidos

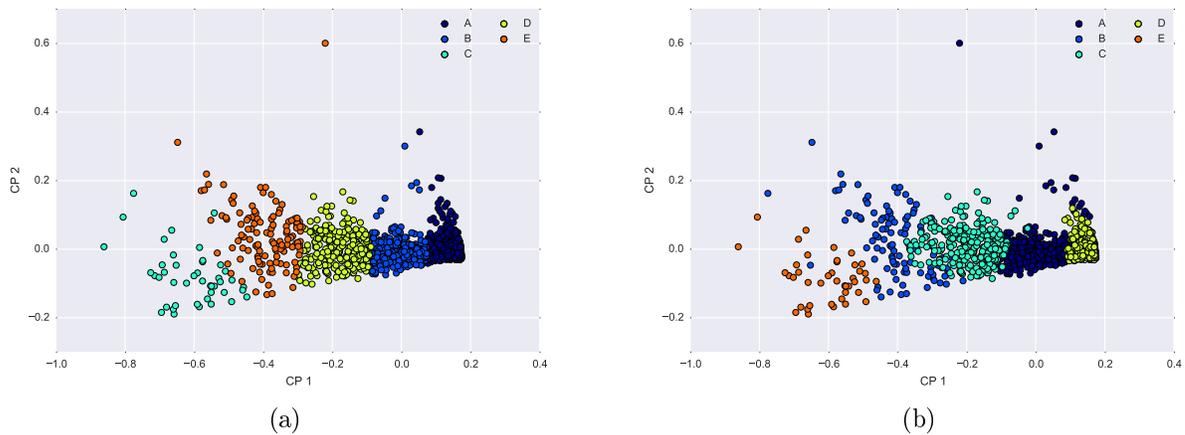


Figura 5.26: 5 clusters definidos por cada método (a) k-Means. (b) Ward

5.5.2.2. Litologías

Al añadir un cluster se puede apreciar un cambio importante en las proporciones de litologías por dominio con respecto a los resultados obtenidos al definir 4 clusters (Figura 5.23).

Se observa que, al igual que en los resultados obtenidos con 4 clusters, existe un dominio (A en k-Means y D en Ward) que agrupa notoriamente la Andesita (AN) y en muy baja cantidad Pórfido Cuprífero (PC), inclusive en mayor grado que en el ejemplo con 4 clusters. Al igual que también se define otro cluster (C en k-Means y E en Ward), que presenta una gran proporción de Pórfido Cuarzófero con respecto a la Andesita. Sin embargo, no logra mayor diferenciación en el Pórfido Feldespático, en los dos casos.

En cuanto a los dominios restantes, se obtiene que, sí bien no presentan variaciones en las proporciones de Pórfido Feldespático, sí lo son con respecto al resto de las litologías, sobre todo en los resultados obtenidos a través de Ward. Sin embargo, queda a criterio del evaluador decidir si estos dominios restantes son los suficientemente disímiles entre sí.

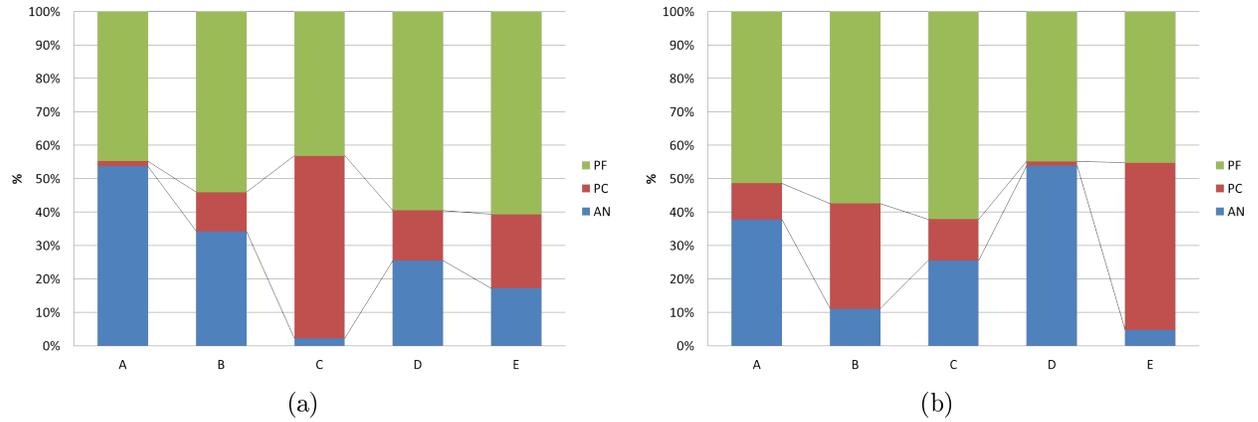


Figura 5.27: Proporciones de litologías por tipo de método para 5 clusters. (a) k-Means. (b) Ward

5.5.2.3. Alteraciones

Se obtienen resultados prácticamente idénticos al caso de 4 clusters, donde se siguen observando en los dos métodos la generación de un cluster (A en k-Means y D en Ward) que agrupa la totalidad de la alteración Clorita-Sericita-Cuarzo (CSC) y la alta mayoría de la alteración Clorita-Sericita-Arcillas (CSA). El resto de los grupos son muy parecidos entre sí y agrupan las alteraciones Sericíticas, por lo que dentro de un determinado criterio pueden ser considerados redundantes y ser unidos dentro de un solo grupo o dominio.

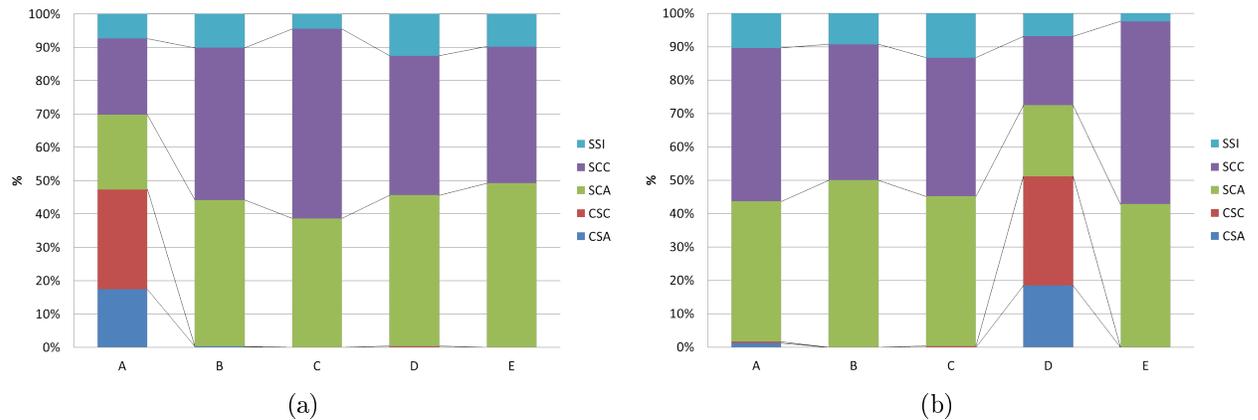


Figura 5.28: Proporciones de alteraciones por tipo de método para 5 clusters. (a) k-Means. (b) Ward

5.5.2.4. Mineralizaciones

Con respecto a las mineralizaciones, los resultados de los dos métodos son semejantes entre sí. Se genera un par de clusters iguales entre si (B y D en k-Means y A y C en Ward), con

proporciones equitativas de los tres tipos de mineralización. Por otro lado, uno de los grupos obtenidos (A en k-Means y D en Ward), presenta una mayor proporción de sulfuros primarios (SP) y agrupa pocos minerales de alto enriquecimiento (HE), junto con otro cluster (C en k-Means y B en Ward) que concentra mayor proporción de minerales asociados a la zona de alto enriquecimiento (HE), siendo este más notorio en el algoritmo Ward. Finalmente, se observa que los minerales de bajo enriquecimiento no son particularmente diferenciados por ningún cluster, salvo quizás por el cluster C de k-Means.

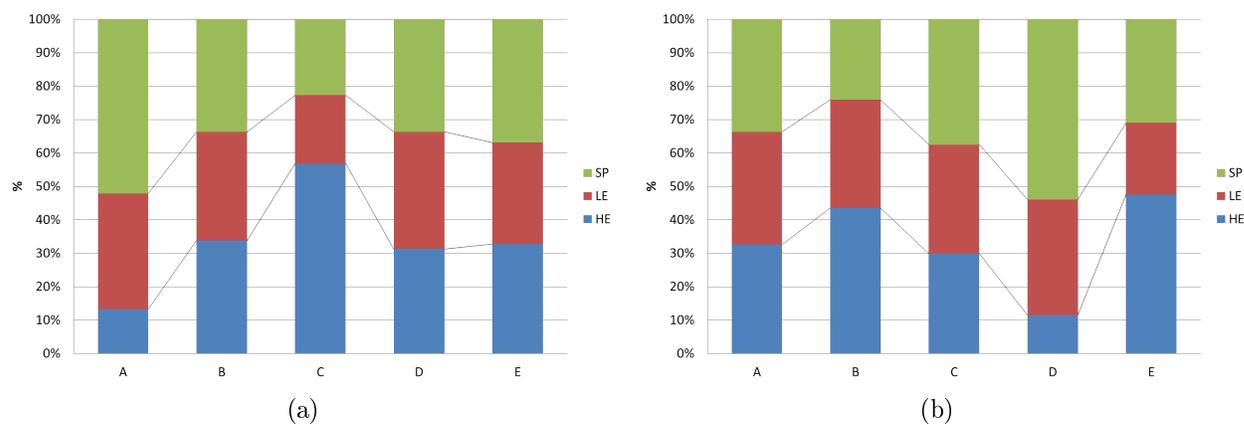


Figura 5.29: Proporciones de mineralizaciones por tipo de método para 5 clusters. (a) k-Means (b) Ward

5.5.3. 6 clusters

5.5.3.1. Clusters definidos

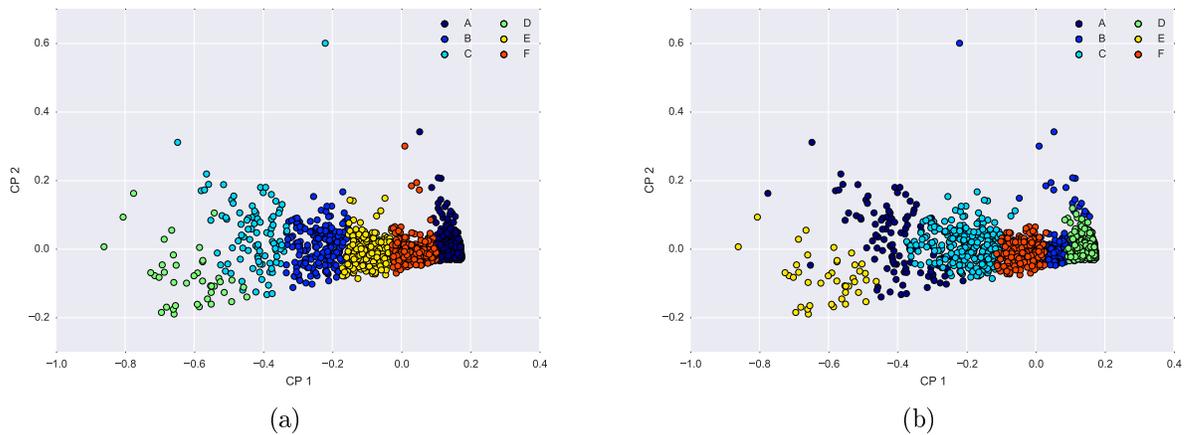


Figura 5.30: 6 clusters definidos por cada método (a) k-Means. (b) Ward

5.5.3.2. Litologías

Al añadir un cluster se puede apreciar un cambio importante en las proporciones de litologías por dominio con respecto a los resultados obtenidos al definir 4 clusters (Figura 5.23).

Siguiendo la tendencia presente al calcular 4 y 5 clusters, también existe un dominio único (A en k-Means y D en Ward) que agrupa notoriamente la Andesita (AN) y en muy baja cantidad Pórfido Cuprífero (PC). Además se genera otro cluster (D en k-Means y E en Ward), que presenta una gran proporción de Pórfido Cuarzífero con respecto a la Andesita. También se destaca que estos dos dominios presentan la misma proporción de Pórfido Feldespático.

Se observa en k-Means (Figura 5.31. (a)) que en los clusters B, C, E y F la proporción de Pórfido Feldespático no varía considerablemente, inclusive los clusters E y F resultan ser muy semejantes entre sí, por lo que pueden ser considerados como redundantes. El cluster C agrupa más datos de Pórfido Cuprífero que el B, dando a entender la diferencia entre ellos.

En cuanto a los dominios obtenidos por Ward (Figura 5.31. (b)), se observa más heterogeneidad en las proporciones obtenidas de los clusters A, B, C y F. Específicamente, en los clusters B, C y F, que presentan porcentajes semejantes de Pórfido Cuarzífero pero difieren en su contenido de Pórfido Feldespático y Andesita, por lo que la decisión de considerar estos dominios como fidedignos dependerá del criterio del investigador a cargo.

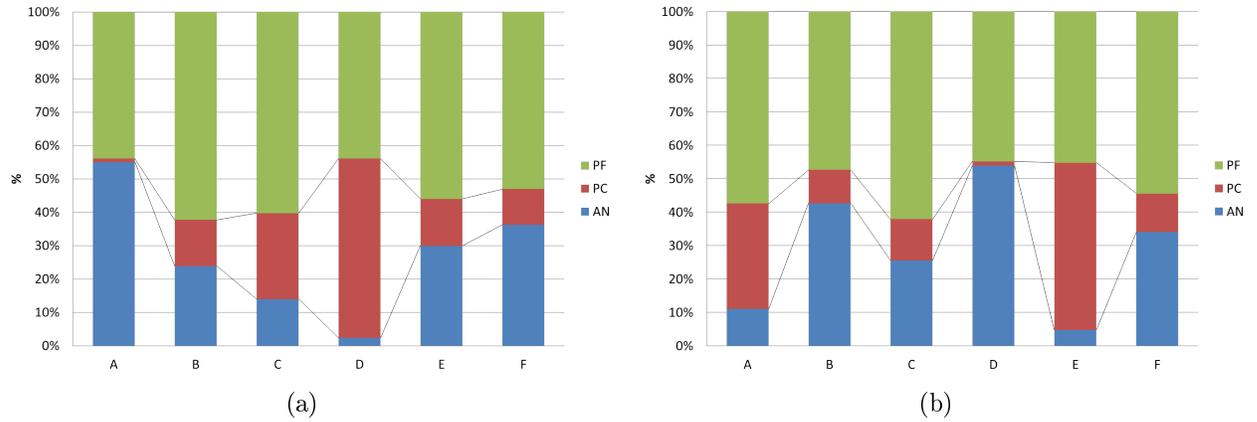


Figura 5.31: Proporciones de litologías por tipo de método para 6 clusters. (a) k-Means. (b) Ward

5.5.3.3. Alteraciones

Se obtienen los mismos resultados que para el caso de 4 y 5 clusters: La definición de un cluster (A en k-Means y D en Ward) que agrupa la totalidad de la alteración Clorita-Sericita-Cuarzo (CSC) y un alto porcentaje de la alteración Clorita-Sericita-Arcillas (CSA). El resto de los grupos siguen siendo muy parecidos entre sí y agrupan las alteraciones Sericíticas, por lo que pueden ser considerados redundantes.

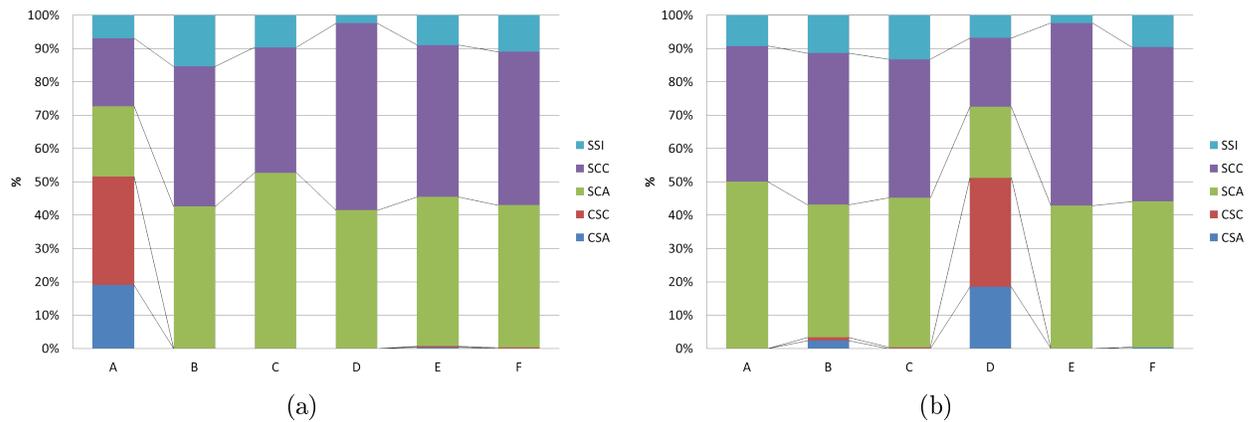


Figura 5.32: Proporciones de alteraciones por tipo de método para 6 clusters. (a) k-Means. (b) Ward

5.5.3.4. Mineralizaciones

En cuanto a los resultados obtenidos a través de k-Means (Figura 5.33. (a)), se destaca que los grupos B, C, E y F presentan proporciones muy semejantes, por lo que dependiendo del criterio aplicado pueden ser considerados como un único dominio. Por otro lado, se observa

que el dominio A presenta un gran porcentaje de Sulfuros Primarios, bajo en minerales de Alto Enriquecimiento (HE) y proporción de Bajo Enriquecimiento prácticamente igual que el resto de los clusters salvo el D. Finalmente, se observa que el cluster D separa minerales de Alto Enriquecimiento y en menor medida los Sulfuros Primarios y minerales de Bajo Enriquecimiento.

Con respecto a los resultados de Ward, los clusters B y C son prácticamente iguales, por lo tanto pueden ser considerados redundantes. Los clusters E y F no distinguen Sulfuros Primarios pero establecen diferencias entre minerales de Alto y Bajo Enriquecimiento. Finalmente, los clusters A y D, si bien presentan proporciones semejantes de minerales de Bajo Enriquecimiento, se diferencian en sus contenidos respectivos de minerales de Alto Enriquecimiento y Sulfuros Primarios.

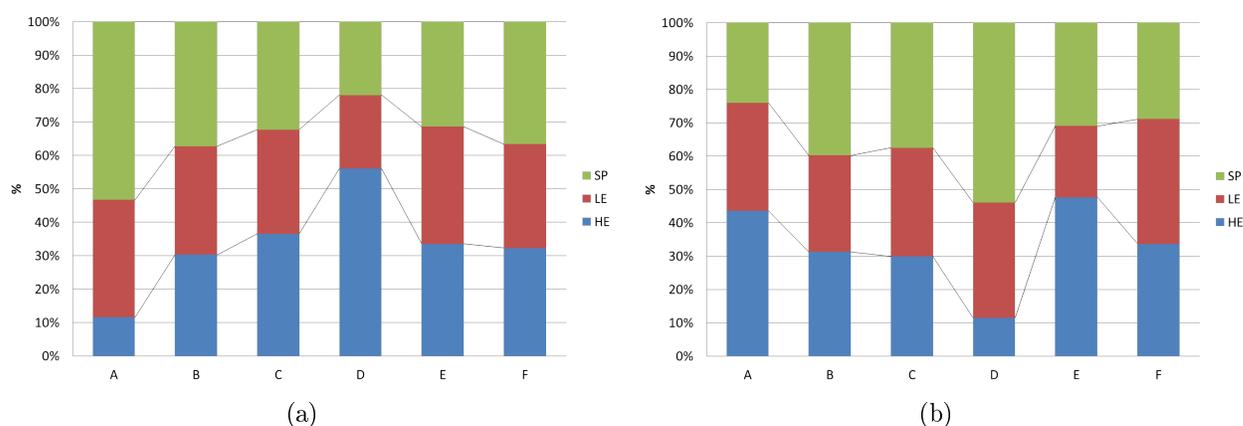


Figura 5.33: Proporciones de mineralizaciones por tipo de método para 6 clusters. (a) k-Means. (b) Ward

5.5.4. Discusiones

En cuanto a las litologías, se observa que en los tres casos, e independiente del algoritmo, se definen dos dominios importantes: uno que concentra un gran porcentaje del Pórfido Cuarzófero (PC) con una baja presencia de Andesita (AN) y otro dominio que, al contrario, presenta un alta presencia de Andesita y una muy baja de Pórfido Cuarzófero. El resto de los dominios van variando dependiendo de la cantidad de clusters definidos. En el caso de 4 clusters, el resto de los dominios no distinguen los datos asociados al Pórfido Cuarzófero, sin embargo, logra separar la Andesita y el Pórfido Feldespático. En el caso de 5 clusters, el agregar un dominio ayuda a clasificar mejor el Pórfido Cuarzófero, en desmedro de una clasificación menos fina del Pórfido Feldespático. En cuanto al caso de 6 clusters, se obtienen resultados semejantes al caso de 5, donde inclusive ya se pueden observar la presencia de clusters muy semejantes en sus proporciones.

Con respecto a las alteraciones, se concluye en todos los casos, independiente del algoritmo usado, que las alteraciones Clorita-Sericita-Cuarzo (CSC) y Clorita-Sericita-Arcillas (CSA) pertenecen a un único dominio y el resto de los dominios presentan proporciones semejantes

de las alteraciones, por lo tanto, definir más de dos dominios no aporta mayor clasificación con respecto a las alteraciones.

En el caso de las mineralizaciones, independiente del modelo, se generan 3 tipos de dominios: uno que presenta un alto porcentaje de mineralización de Alto Enriquecimiento (HE) y baja proporción de mineralización de Bajo Enriquecimiento (LE). Un segundo dominio, que al contrario del primero, presenta una alta proporción de mineralización LE y una baja de HE, y finalmente un tercer tipo de dominio en donde se tienen proporciones aproximadamente equitativas entre las mineralizaciones HE y LE, junto con los Sulfuros Primarios. Se debe destacar también que ninguno de estos tres tipos de dominios presenta preferencia en la clasificación de los Sulfuros Primarios.

Finalmente, se concluye, que si bien existen dominios que agrupan tajantemente variables geológicas, queda al criterio del investigador decidir si los clusters que presentan proporciones semejantes entre sí deben ser unidos, omitidos, o pueden ser clasificados como grupos o dominios separados.

Capítulo 6

Conclusiones y recomendaciones

Se logra establecer una metodología para determinar dominios geometalúrgicos, la cual muestra un potencial interesante ya que permite generar estas unidades de forma no supervisada o con poca intervención del investigador a cargo de analizar los datos. Sin embargo, cuando se requiere esta intervención, es necesario visualizar los resultados y en base al criterio del analista se deben tomar decisiones en cuanto a si los dominios obtenidos de forma automática presentan sentido o no.

Al ser el primer paso en esta metodología, la selección de variables relevantes se presenta como una tarea delicada, ya que por un lado pueden existir variables que no aporten mayor variabilidad a los datos. Sin embargo, éstas pueden estar correlacionadas con la variable geometalúrgica a estudiar. Si bien en esta metodología se optó por utilizar el coeficiente de correlación de Pearson entre las variables de geoquímica y sintéticas con respecto a la recuperación de cobre, los valores usados como criterio surgen de la simple selección de las correlaciones más altas, tanto positivas como negativas, por lo que se recomienda realizar un análisis de sensibilidad en ese sentido para darle más rigurosidad al estudio. Otra alternativa es dejar la selección de las variables relevantes en manos del modelador, eligiendo las variables que estime convenientes, que posean tanto sentido geológico como relación con la variable geometalúrgica a estudiar.

El estudio de las componentes principales seleccionadas permitió apreciar características no observadas explícitamente a partir de los datos. Por ejemplo, al analizar la primera componente principal, se obtuvo que a menor valor de esta el contenido de arcillas es mayor, lo que posteriormente permitió identificar clusters en los que se proyecta una menor recuperación de cobre.

En cuanto a la generación de los clusters en el caso base, se observó que al definir 4 dominios no existe mayor diferencia en los resultados obtenidos por cada modelo, por lo que la elección entre k-Means y Ward dependerá de un criterio a elección. El criterio elegido fue el coeficiente de silueta, que dimensiona cuan bien se encuentran agrupados los datos. Se obtuvo que k-Means presentaba mayor coeficiente de silueta, por lo que se prefirieron sus resultados en vez de los obtenidos por Ward.

Se puede apreciar en los resultados de los modelos predictivos que, en general, el coeficiente de determinación (R^2) es relativamente bajo para los dominios definidos por los dos métodos indistintamente. En cuanto al algoritmo utilizado para realizar las regresiones multi-lineales (LARS Lasso), es particularmente útil la existencia de un factor de sensibilidad, que permite realizar un *trade-off* entre la cantidad de variables y el ajuste del modelo. Sin embargo, los resultados indican que la relación entre las variables de geoquímica y sintéticas con respecto a la recuperación de cobre es no lineal, por lo que es recomendable realizar el modelo predictivo en base a alternativas, tales como redes neuronales o árboles de decisión.

En cuanto al análisis de sensibilidad, se observó que cada variable geológica debe ser analizada independientemente. En el caso de las litologías, independiente del número de clusters definidos, se definen dos dominios que son diferenciados por sus proporciones de Pórfido Cuarzífero (PC) y Andesita (AN) respectivamente. También se destaca que con la presencia de 6 clusters ya se identifican dominios con proporciones muy semejantes entre ellos, por lo que dependiendo del criterio pueden ser considerados unidades redundantes.

Con respecto a las alteraciones se concluyó que con al menos dos clusters ya se pueden clasificar los datos, generando un dominio que agrupa parte de las alteraciones Sericita-Cuarzo y la totalidad de las alteraciones Clorita-Sericita-Cuarzo (CSC) y Clorita-Sericita-Arcillas (CSA), junto con otro dominio que agrupa el resto de los datos de Sericita-Cuarzo. El añadir más dominios generará redundancia y no aportará mayor clasificación a las alteraciones.

En el caso de las mineralizaciones, independiente del modelo, se generaron tres tipos de dominios: uno con alto porcentaje de mineralización de alto enriquecimiento (HE) y baja proporción de bajo enriquecimiento (LE), otro que al contrario presenta alto LE y bajo HE. El tercer tipo de dominio presenta distribuciones relativamente equitativas entre sus contenidos de HE, LE y Sulfuros Primarios (SP). Añadir más clusters genera dominios que presentan características semejantes al tercer tipo de dominio obtenido, por lo que a partir de este número resulta redundante la adición de más clusters de clasificación.

Como comentario final respecto al análisis de sensibilidad, es labor del investigador revisar las proporciones de las variables geológicas por cada dominio obtenido, independiente del método utilizado, con la finalidad de determinar el número óptimo de clusters que permita clasificar correctamente los dominios, sin caer en la redundancia.

Finalmente, una vez definidos los dominios a utilizar, su aplicación en datos nuevos puede ser realizada en base a las proporciones de variables geológicas asociadas a cada dominio.

Bibliografía

- [1] BHP Billiton. Marco geológico Minera Escondida. Informe interno, 2009.
- [2] J. Boisvert, M. Rossi, K. Ehrig, and C. Deutsch. Geometallurgical modeling at Olympic Dam mine, South Australia. *Mathematical Geosciences*, pages 1–25, 2013.
- [3] S. Bulatovic, D. Wyslouzil, and C. Kant. Effect of clay slimes on copper, molybdenum flotation from porphyry ores. *Proceedings of Cooper*, 2:95–109, 1999.
- [4] G. Corbett and T. Leach. Southwest pacific rim gold-copper system: Structure, alteration, and mineralization. *Economic Geology*, page 238, 1998.
- [5] S. Coward, J. Vann, S. Dunham, and M. Stewart. The primary-response framework for geometallurgical variables. In *Seventh International Mining Geology Conference*, pages 109–113, Perth, WA., 2009.
- [6] S. Dunham, J. Vann, and S. Coward. Beyond geometallurgy: gaining competitive advantage by exploiting the broad view of geometallurgy. In *The first AusIMM international geometallurgy conference*, 2011.
- [7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):pp. 407–451, 2004.
- [8] X. Emery. Fundamentos de geoestadística. Apuntes del curso de Simulación Geoestadística (MI6041) de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.
- [9] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 26:417–441, 1933.
- [10] A.J. Izenman. *Modern Multivariate Statistical Techniques*. Springer, 2008.
- [11] C. Jeraldo, F. Ramírez, and C. Sáez. Redefinition of geo-metallurgical Wi model using k-means clustering. In *XIII Congreso Geológico Chileno*, pages 936–938. Universidad Católica del Norte, Agosto 2012.
- [12] W. Kracht. Apuntes del curso de ingeniería del procesamiento de minerales. Departamento de Ingeniería de Minas. Facultad de Ciencias Físicas y Matemáticas. Universidad de Chile., 2011.

- [13] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, page 14. California, USA, 1967.
- [14] V. Maksaev. Apuntes del curso de metalogénesis. Departamento de Geología. Facultad de Ciencias Físicas y Matemáticas. Universidad de Chile.
- [15] R. A. Padilla, S. R. Titley, and F. Pimentel. Geology of the Escondida porphyry copper deposit. *Economic Geology*, 96:307–324, 2001.
- [16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [17] F. Provost, T. Fawcett, and R. Kohavi. The case against accuracy estimation for comparing induction algorithms. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 445–453. Morgan Kaufmann, 1997.
- [18] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53 – 65, 1987.
- [19] R. Sillitoe. The tops and bottoms of porphyry copper deposits. *Economic Geology*, 68:799–815, 1973.
- [20] R.H. Sillitoe. Porphyry copper systems. *Society of Economic Geologists*, 105:3–41, 2010.
- [21] B. Townley, M. Muñoz, and R. Luca. Modelamiento geoquímico distrito Escondida: discriminación de ambientes hidrotermales. Technical report, GeoAV Geochemistry & Exploration S.A., 2011.
- [22] W. O. Véliz. Relación espacio-temporal del sistema pórfido cuprífero y epitermal en el yacimiento Escondida. Master’s thesis, Universidad Católica del Norte, 2004.
- [23] J.H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

Anexo A

Estudio exploratorio de datos

A.1. Base de datos original

En la Tabla A.1 se encuentran las estadísticas de las variables geoquímicas para la base de datos original, es decir, sin la remoción de variables geológicas con pocos datos y/o irrelevantes.

Tabla A.1: Estadísticas de leyes

Elemento	N. de datos	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
Ag	1916	3.47	3.44	0.99	0.04	1.47	2.68	4.56	51.30	3.09
Al	1916	0.84	0.60	0.71	0.08	0.42	0.60	1.04	3.81	0.62
As	1916	46.96	206.59	4.40	0.10	3.70	10.90	31.58	6610.00	27.88
Au	979	0.21	0.09	0.42	0.20	0.20	0.20	0.20	1.90	0.00
B	1916	10.17	1.44	0.14	10.00	10.00	10.00	10.00	40.00	0.00
Ba	1916	37.28	19.76	0.53	4.80	30.00	30.00	40.00	240.00	10.00
Be	1916	0.25	0.23	0.92	0.05	0.08	0.15	0.36	1.56	0.28
Bi	1916	1.32	1.70	1.28	0.03	0.35	0.82	1.65	19.85	1.30
Ca	1916	0.14	0.22	1.55	0.01	0.02	0.07	0.19	3.89	0.17
Cd	1916	4.13	12.97	3.14	0.01	0.29	1.11	3.44	345.00	3.15
Ce	1916	17.86	9.07	0.51	0.69	10.70	16.70	24.60	45.50	13.90
Co	1916	12.38	11.90	0.96	0.80	6.60	10.70	15.70	349.00	9.10
Cr	1916	115.01	113.37	0.99	1.00	37.00	80.00	164.00	1560.00	127.00
Cs	1916	1.09	0.58	0.53	0.13	0.69	0.94	1.36	5.84	0.67
Cu	1916	0.70	0.23	0.33	0.17	0.52	0.69	0.93	1.00	0.41
Fe	1916	2.32	1.32	0.57	0.00	1.50	2.04	2.83	13.58	1.33
Ga	1916	2.32	2.03	0.87	0.25	0.89	1.39	3.05	11.35	2.16
Ge	1916	0.06	0.03	0.42	0.05	0.05	0.05	0.07	0.58	0.02
Hf	1916	0.02	0.00	0.18	0.02	0.02	0.02	0.02	0.09	0.00
Hg	1916	0.03	0.07	2.30	0.01	0.01	0.01	0.03	1.44	0.02
In	1916	0.32	0.50	1.54	0.01	0.08	0.16	0.37	9.07	0.29
K	1916	0.22	0.09	0.39	0.01	0.17	0.20	0.25	0.76	0.08
La	1916	8.88	4.44	0.50	0.30	5.50	8.40	12.20	23.90	6.70
Li	1916	4.62	6.73	1.46	0.00	1.00	1.00	6.00	33.00	5.00
Mg	1916	0.37	0.56	1.51	0.01	0.03	0.04	0.51	2.42	0.48
Mn	1916	0.02	0.04	1.72	0.00	0.00	0.00	0.03	0.29	0.02
Mo	1916	92.82	95.46	1.03	0.98	35.90	63.00	119.00	1085.00	83.10
Na	1916	0.06	0.03	0.46	0.01	0.04	0.05	0.07	0.21	0.03
Nb	1916	0.05	0.02	0.28	0.05	0.05	0.05	0.05	0.23	0.00
Ni	1916	14.19	17.54	1.24	1.40	4.40	7.20	14.20	200.30	9.80
P	1916	393.65	452.20	1.15	10.00	70.00	170.00	580.00	2910.00	510.00
Pb	1916	57.60	171.11	2.97	0.30	4.30	19.60	58.38	4550.00	54.08
Rb	1916	12.49	6.39	0.51	0.60	9.00	11.00	13.90	60.30	4.90
Re	1916	0.73	1.44	1.98	0.00	0.14	0.35	0.70	25.50	0.56
S	1916	2.06	1.40	0.68	0.14	1.20	1.75	2.47	10.00	1.27

Tabla A.1: Estadísticas de leyes

Elemento	N. de datos	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
Sb	1916	1.94	10.30	5.32	0.05	0.13	0.26	0.85	306.00	0.72
Sc	1916	1.45	1.68	1.15	0.10	0.50	0.70	1.80	13.30	1.30
Se	1916	3.82	1.65	0.43	0.20	2.70	3.50	4.60	15.50	1.90
Sn	1916	0.75	1.67	2.23	0.20	0.30	0.50	0.70	59.00	0.40
Sr	1916	102.31	128.23	1.25	4.40	31.15	60.90	118.13	1240.00	86.98
Ta	1916	0.01	0.00	0.24	0.01	0.01	0.01	0.01	0.07	0.00
Te	1916	0.70	1.48	2.12	0.03	0.21	0.37	0.66	26.70	0.45
Th	1916	1.80	0.98	0.55	0.20	1.00	1.60	2.50	6.10	1.50
Ti	1916	0.01	0.02	1.86	0.01	0.01	0.01	0.01	0.20	0.00
Tl	1916	0.18	0.12	0.66	0.02	0.11	0.15	0.21	1.97	0.10
U	1916	0.43	0.30	0.70	0.05	0.25	0.36	0.51	3.28	0.26
V	1916	14.80	20.67	1.40	1.00	3.00	5.00	17.00	168.00	14.00
W	1916	0.57	1.83	3.22	0.05	0.18	0.32	0.56	50.10	0.38
Y	1916	4.25	5.58	1.31	0.06	0.51	0.99	7.38	48.00	6.87
Zn	1916	516.69	854.81	1.65	2.00	24.00	187.50	740.50	10000.00	716.50
Zr	1916	1.00	0.02	0.02	1.00	1.00	1.00	1.00	2.00	0.00

Tabla A.2: Estadísticas de recuperación

Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
84.92 %	7.50 %	0.09	36.47 %	81.41 %	86.71 %	90.32 %	98.14 %	8.91 %

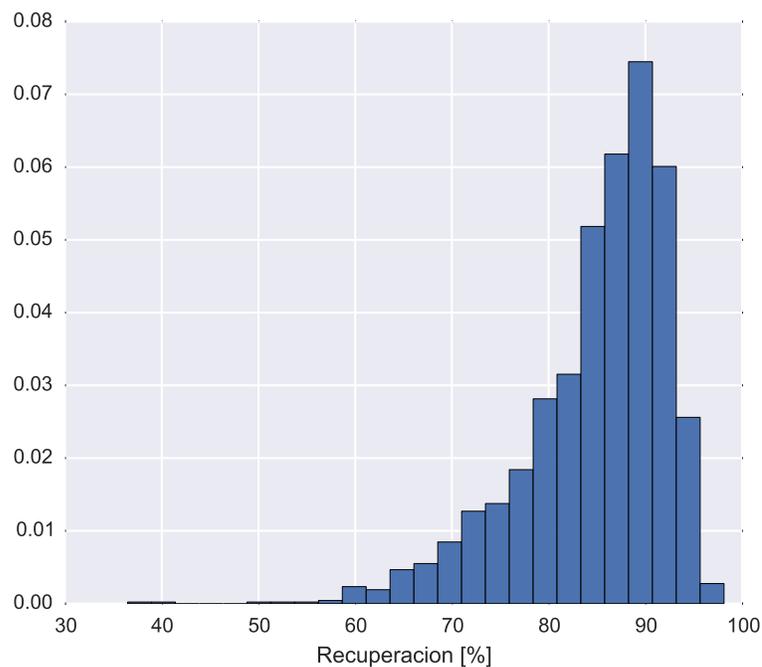


Figura A.1: Histograma de la recuperación de cobre a partir de los datos originales

Tabla A.3: Estadísticas de variables sintéticas en base de datos original.

Elemento	Media	Desv. Est.	CV	Mínimo	25 %	50 %	75 %	Máximo	RIC
cu/s	0.46	0.32	0.69	0.04	0.24	0.40	0.61	7.14	0.37
k/al	0.34	0.14	0.43	0.05	0.22	0.33	0.44	0.70	0.22
kxal	0.21	0.24	1.11	0.00	0.08	0.12	0.25	1.92	0.17
(kxna)/al	0.02	0.01	0.41	0.00	0.01	0.02	0.02	0.10	0.01
na/al	0.08	0.04	0.51	0.01	0.05	0.08	0.10	1.00	0.05
(al+k)/(na+ca+mg)	4.36	3.21	0.74	0.13	1.44	3.58	6.67	17.78	5.22
(al+k+na)/(ca+mg)	7.62	6.85	0.90	0.13	1.64	4.94	12.67	37.50	11.02
(ca+na)/(k+al)	0.21	0.28	1.34	0.03	0.10	0.14	0.23	7.80	0.14
al/(na+ca+k)	1.90	1.34	0.71	0.05	0.86	1.63	2.60	10.50	1.74
al/mg	11.86	9.49	0.80	0.24	1.81	12.50	18.50	48.00	16.69
3*al/(na+k)	8.78	4.75	0.54	1.85	5.57	7.20	10.27	35.67	4.70
(k+al+s)/(fe+s)	0.75	0.14	0.19	0.34	0.66	0.73	0.82	1.91	0.16
k/mg	4.68	4.32	0.92	0.09	0.40	4.60	7.33	27.00	6.93
mg/al	0.28	0.32	1.13	0.02	0.05	0.08	0.55	4.19	0.50
mn/al	0.02	0.03	1.40	0.00	0.00	0.01	0.02	0.32	0.02

A.2. Base de datos final

A.2.1. Matriz de correlación

En esta sección se detalla la matriz de correlación de todas las variables continuas presentes en la base de datos final. Como esta matriz presenta un gran tamaño, se subdividió en 7 tablas:

1. Desde $(Al + K)/(Na + Ca + Mg)$ hasta $(K + Al + S)/(Fe + S)$ (página 94)
2. Desde $(KxNa)/Al$ hasta Au (página 95)
3. Desde B hasta Cu (página 97)
4. Desde Cu/S hasta $KxAl$ (página 98)
5. Desde La hasta Ni (página 100)
6. Desde P hasta Ta (página 101)
7. Desde Te hasta Zr (página 103)

Tabla A.4: Matriz de correlación. Parte 1 de 7

	$(al+k)/(na+ca+mg)$	$(al+k+na)/(ca+mg)$	$(ca+na)/(k+al)$	$(k+al+s)/(fe+s)$
$(al+k)/(na+ca+mg)$	1.00	0.96	-0.37	0.10
$(al+k+na)/(ca+mg)$	0.96	1.00	-0.34	0.10
$(ca+na)/(k+al)$	-0.37	-0.34	1.00	-0.16
$(k+al+s)/(fe+s)$	0.10	0.10	-0.16	1.00
$(kxna)/al$	-0.08	0.00	0.00	0.06
$3*al/(na+k)$	-0.35	-0.33	-0.11	0.26
ag	-0.03	-0.02	0.01	-0.05
al	-0.39	-0.38	-0.11	0.37
$al/(na+ca+k)$	0.88	0.79	-0.40	-0.01
al/mg	0.75	0.74	-0.09	-0.03
as	0.00	0.00	0.14	-0.12

Tabla A.4: Matriz de correlación. Parte 1 de 7

	$(al+k)/(na+ca+mg)$	$(al+k+na)/(ca+mg)$	$(ca+na)/(k+al)$	$(k+al+s)/(fe+s)$
au	0.03	0.02	-0.05	-0.06
b	-0.02	-0.03	0.06	-0.04
ba	-0.02	-0.04	-0.06	0.30
be	-0.52	-0.51	-0.05	0.28
bi	0.00	0.02	0.09	-0.19
ca	-0.50	-0.48	0.66	0.05
cd	-0.02	-0.01	-0.04	0.03
ce	-0.37	-0.35	-0.13	0.33
co	0.12	0.12	-0.08	-0.17
cr	-0.07	-0.05	0.08	-0.10
cs	-0.28	-0.31	-0.06	0.11
cu	0.11	0.11	0.04	0.09
cu/s	-0.11	-0.08	-0.11	0.53
fe	-0.23	-0.24	0.22	-0.49
ga	-0.53	-0.51	-0.05	0.28
ge	-0.25	-0.24	0.15	-0.04
hf	-0.12	-0.12	-0.01	0.08
hg	0.02	0.01	0.02	-0.11
in	0.04	0.06	-0.05	-0.08
k	-0.10	-0.14	-0.13	0.38
k/al	0.32	0.27	0.15	-0.36
k/mg	0.59	0.55	0.00	-0.22
kxal	-0.35	-0.34	-0.07	0.36
la	-0.33	-0.30	-0.14	0.33
li	-0.54	-0.51	-0.01	0.22
mg	-0.57	-0.54	0.00	0.21
mg/al	-0.66	-0.62	0.07	0.13
mn	-0.50	-0.47	0.04	0.11
mn/al	-0.52	-0.48	0.16	0.00
mo	0.00	0.03	-0.02	0.04
na	-0.45	-0.36	-0.09	0.36
na/al	-0.07	0.03	0.22	-0.25
nb	-0.13	-0.12	0.08	-0.06
ni	-0.10	-0.09	0.06	-0.05
p	-0.55	-0.53	0.05	0.21
pb	-0.12	-0.11	0.00	0.04
rb	-0.18	-0.20	-0.12	0.34
re	0.21	0.24	-0.11	-0.05
s	0.08	0.05	0.28	-0.49
sb	-0.04	-0.04	0.18	-0.10
sc	-0.43	-0.41	-0.05	0.28
se	-0.01	0.03	0.05	-0.02
sn	-0.06	-0.06	0.26	-0.19
sr	-0.13	-0.16	0.00	0.13
ta	-0.01	-0.02	-0.01	0.04
te	-0.06	-0.06	0.32	-0.23

Tabla A.4: Matriz de correlación. Parte 1 de 7

	$(al+k)/(na+ca+mg)$	$(al+k+na)/(ca+mg)$	$(ca+na)/(k+al)$	$(k+al+s)/(fe+s)$
th	-0.46	-0.43	-0.07	0.26
ti	-0.24	-0.22	0.01	0.14
tl	0.27	0.26	-0.12	0.03
u	-0.05	-0.03	-0.10	0.06
v	-0.50	-0.47	-0.01	0.21
w	0.00	-0.02	0.04	-0.05
y	-0.53	-0.50	-0.01	0.23
zn	-0.30	-0.28	-0.02	0.13
zr	-0.02	-0.02	-0.01	0.02

Tabla A.5: Matriz de correlación. Parte 2 de 7

	$(kxna)/al$	$3*al/(na+k)$	ag	al	$al/(na+ca+k)$	al/mg	as	au
$(al+k)/(na+ca+mg)$	-0.08	-0.35	-0.03	-0.39	0.88	0.75	0.00	0.03
$(al+k+na)/(ca+mg)$	0.00	-0.33	-0.02	-0.38	0.79	0.74	0.00	0.02
$(ca+na)/(k+al)$	0.00	-0.11	0.01	-0.11	-0.40	-0.09	0.14	-0.05
$(k+al+s)/(fe+s)$	0.06	0.26	-0.05	0.37	-0.01	-0.03	-0.12	-0.06
$(kxna)/al$	1.00	-0.46	0.04	-0.18	-0.04	-0.06	-0.05	0.01
$3*al/(na+k)$	-0.46	1.00	-0.07	0.88	-0.36	-0.48	-0.10	-0.02
ag	0.04	-0.07	1.00	-0.09	-0.04	-0.07	0.30	0.21
al	-0.18	0.88	-0.09	1.00	-0.35	-0.53	-0.13	-0.04
$al/(na+ca+k)$	-0.04	-0.36	-0.04	-0.35	1.00	0.50	-0.01	0.07
al/mg	-0.06	-0.48	-0.07	-0.53	0.50	1.00	0.06	-0.03
as	-0.05	-0.10	0.30	-0.13	-0.01	0.06	1.00	0.00
au	0.01	-0.02	0.21	-0.04	0.07	-0.03	0.00	1.00
b	0.23	-0.07	-0.04	-0.05	-0.04	0.01	0.01	-0.02
ba	0.17	0.06	-0.06	0.20	-0.02	-0.11	-0.10	-0.01
be	-0.08	0.70	-0.01	0.80	-0.41	-0.65	-0.10	0.00
bi	0.04	-0.12	0.65	-0.16	0.00	-0.01	0.28	0.18
ca	-0.05	0.32	-0.03	0.44	-0.50	-0.36	-0.04	-0.08
cd	-0.02	0.11	0.15	0.11	-0.03	-0.07	0.02	-0.02
ce	0.11	0.42	0.05	0.53	-0.28	-0.58	-0.15	0.05
co	-0.08	0.09	0.04	0.08	0.12	0.09	0.02	0.09
cr	0.01	-0.10	0.18	-0.16	-0.11	-0.03	0.04	0.22
cs	-0.05	0.42	0.02	0.52	-0.25	-0.35	-0.06	-0.07
cu	0.09	-0.23	0.39	-0.25	0.06	0.09	0.13	0.17
cu/s	0.16	0.10	0.04	0.15	-0.13	-0.23	-0.13	0.05
fe	-0.23	0.27	0.16	0.25	-0.18	-0.17	0.24	-0.01
ga	-0.08	0.78	-0.01	0.92	-0.42	-0.66	-0.11	-0.02
ge	-0.07	0.22	0.22	0.26	-0.21	-0.24	0.48	0.02
hf	-0.01	0.15	0.00	0.20	-0.10	-0.13	-0.03	-0.03
hg	0.00	-0.08	0.41	-0.10	0.03	0.04	0.38	-0.02
in	0.06	-0.04	0.46	-0.04	0.03	-0.04	0.11	0.05
k	0.37	0.16	-0.08	0.56	0.02	-0.23	-0.13	-0.04
k/al	0.34	-0.84	-0.02	-0.76	0.39	0.50	0.07	-0.02
k/mg	0.02	-0.59	-0.11	-0.59	0.49	0.89	0.06	-0.04

Tabla A.5: Matriz de correlación. Parte 2 de 7

	$(kxna)/al$	$3*al/(na+k)$	ag	al	$al/(na+ca+k)$	al/mg	as	au
kxal	0.09	0.60	-0.08	0.89	-0.27	-0.46	-0.10	-0.04
la	0.13	0.37	0.07	0.47	-0.24	-0.55	-0.15	0.06
li	-0.12	0.77	-0.06	0.90	-0.45	-0.64	-0.11	-0.04
mg	-0.11	0.78	-0.06	0.91	-0.45	-0.66	-0.11	-0.04
mg/al	0.05	0.59	0.00	0.68	-0.49	-0.77	-0.10	-0.02
mn	-0.10	0.58	0.07	0.68	-0.42	-0.58	-0.09	-0.05
mn/al	-0.02	0.32	0.16	0.35	-0.42	-0.57	-0.03	-0.02
mo	0.04	0.05	0.00	0.06	-0.01	-0.01	0.01	-0.02
na	0.37	0.52	-0.01	0.73	-0.45	-0.57	-0.12	0.01
na/al	0.61	-0.49	0.10	-0.51	-0.12	0.06	0.16	0.02
nb	0.09	0.02	0.09	0.10	-0.10	-0.10	0.08	0.04
ni	0.01	0.12	0.06	0.15	-0.07	-0.10	-0.02	0.03
p	-0.10	0.72	-0.03	0.83	-0.49	-0.62	-0.09	-0.05
pb	-0.01	0.09	0.39	0.12	-0.10	-0.17	0.08	-0.01
rb	0.28	0.24	-0.02	0.60	-0.08	-0.30	-0.12	-0.06
re	0.03	-0.09	0.15	-0.12	0.15	0.17	0.04	-0.03
s	-0.15	-0.23	0.26	-0.30	0.07	0.23	0.37	-0.01
sb	-0.04	-0.07	0.36	-0.10	-0.04	0.00	0.79	-0.02
sc	0.01	0.62	-0.05	0.82	-0.34	-0.53	-0.10	-0.05
se	0.09	-0.11	0.43	-0.13	-0.04	-0.04	0.15	0.35
sn	-0.05	-0.09	0.17	-0.10	-0.06	0.03	0.22	-0.01
sr	-0.11	0.28	-0.01	0.26	-0.12	-0.12	0.00	0.03
ta	-0.01	0.00	0.01	0.00	-0.01	-0.01	0.05	-0.01
te	-0.03	-0.15	0.28	-0.19	-0.08	0.06	0.39	-0.01
th	0.11	0.42	-0.01	0.50	-0.32	-0.64	-0.16	0.06
ti	0.31	0.14	-0.05	0.42	-0.17	-0.27	-0.05	-0.04
tl	0.07	-0.14	0.16	-0.07	0.21	0.28	0.05	-0.02
u	0.11	0.07	0.32	0.09	-0.03	-0.17	0.04	0.14
v	-0.03	0.69	-0.06	0.88	-0.41	-0.58	-0.09	-0.05
w	0.00	-0.07	0.15	-0.07	0.00	0.04	0.08	0.00
y	-0.02	0.57	0.00	0.70	-0.41	-0.63	-0.10	-0.04
zn	0.00	0.32	0.45	0.39	-0.25	-0.41	0.05	-0.02
zr	-0.01	0.04	0.00	0.06	-0.01	-0.03	0.00	0.00

Tabla A.6: Matriz de correlación. Parte 3 de 7

	b	ba	be	bi	ca	cd	ce	co	cr	cs	cu
$(al+k)/(na+ca+mg)$	-0.02	-0.02	-0.52	0.00	-0.50	-0.02	-0.37	0.12	-0.07	-0.28	0.11
$(al+k+na)/(ca+mg)$	-0.03	-0.04	-0.51	0.02	-0.48	-0.01	-0.35	0.12	-0.05	-0.31	0.11
$(ca+na)/(k+al)$	0.06	-0.06	-0.05	0.09	0.66	-0.04	-0.13	-0.08	0.08	-0.06	0.04
$(k+al+s)/(fe+s)$	-0.04	0.30	0.28	-0.19	0.05	0.03	0.33	-0.17	-0.10	0.11	0.09
$(kxna)/al$	0.23	0.17	-0.08	0.04	-0.05	-0.02	0.11	-0.08	0.01	-0.05	0.09
$3*al/(na+k)$	-0.07	0.06	0.70	-0.12	0.32	0.11	0.42	0.09	-0.10	0.42	-0.23
ag	-0.04	-0.06	-0.01	0.65	-0.03	0.15	0.05	0.04	0.18	0.02	0.39
al	-0.05	0.20	0.80	-0.16	0.44	0.11	0.53	0.08	-0.16	0.52	-0.25
$al/(na+ca+k)$	-0.04	-0.02	-0.41	0.00	-0.50	-0.03	-0.28	0.12	-0.11	-0.25	0.06

Tabla A.6: Matriz de correlación. Parte 3 de 7

	b	ba	be	bi	ca	cd	ce	co	cr	cs	cu
al/mg	0.01	-0.11	-0.65	-0.01	-0.36	-0.07	-0.58	0.09	-0.03	-0.35	0.09
as	0.01	-0.10	-0.10	0.28	-0.04	0.02	-0.15	0.02	0.04	-0.06	0.13
au	-0.02	-0.01	0.00	0.18	-0.08	-0.02	0.05	0.09	0.22	-0.07	0.17
b	1.00	0.04	-0.05	0.06	-0.01	-0.03	-0.09	-0.02	0.01	0.02	0.06
ba	0.04	1.00	0.19	-0.13	0.10	-0.01	0.22	-0.10	0.00	0.07	-0.02
be	-0.05	0.19	1.00	-0.10	0.38	0.07	0.62	-0.02	-0.09	0.49	-0.17
bi	0.06	-0.13	-0.10	1.00	-0.05	0.08	-0.02	0.11	0.14	-0.02	0.33
ca	-0.01	0.10	0.38	-0.05	1.00	0.04	0.22	-0.01	-0.04	0.28	-0.11
cd	-0.03	-0.01	0.07	0.08	0.04	1.00	0.10	0.07	0.01	0.12	0.07
ce	-0.09	0.22	0.62	-0.02	0.22	0.10	1.00	-0.11	-0.03	0.36	-0.11
co	-0.02	-0.10	-0.02	0.11	-0.01	0.07	-0.11	1.00	-0.02	0.03	-0.10
cr	0.01	0.00	-0.09	0.14	-0.04	0.01	-0.03	-0.02	1.00	-0.15	0.19
cs	0.02	0.07	0.49	-0.02	0.28	0.12	0.36	0.03	-0.15	1.00	-0.23
cu	0.06	-0.02	-0.17	0.33	-0.11	0.07	-0.11	-0.10	0.19	-0.23	1.00
cu/s	-0.01	0.19	0.19	-0.07	-0.01	0.04	0.33	-0.28	0.11	-0.04	0.44
fe	0.03	-0.18	0.17	0.31	0.22	0.03	-0.14	0.23	-0.03	0.21	-0.11
ga	-0.04	0.18	0.80	-0.10	0.47	0.09	0.60	0.02	-0.09	0.48	-0.23
ge	0.00	-0.04	0.28	0.24	0.19	0.01	0.16	0.04	0.04	0.13	0.07
hf	0.06	0.03	0.20	-0.03	0.09	0.01	0.12	0.01	-0.04	0.13	-0.02
hg	0.01	-0.07	-0.06	0.24	-0.06	0.09	-0.09	0.05	-0.01	-0.01	0.12
in	-0.01	-0.06	-0.01	0.42	-0.04	0.25	0.11	0.15	0.07	0.10	0.16
k	-0.01	0.38	0.43	-0.17	0.31	0.04	0.37	0.02	-0.20	0.40	-0.15
k/al	0.08	-0.09	-0.66	0.04	-0.22	-0.12	-0.53	-0.01	0.04	-0.39	0.12
k/mg	0.01	-0.15	-0.65	-0.04	-0.28	-0.10	-0.62	0.09	-0.05	-0.38	0.03
kxal	-0.04	0.27	0.70	-0.16	0.46	0.08	0.46	0.07	-0.17	0.51	-0.23
la	-0.10	0.23	0.58	0.00	0.18	0.10	0.99	-0.12	-0.02	0.33	-0.09
li	-0.03	0.13	0.78	-0.12	0.49	0.09	0.58	0.04	-0.10	0.51	-0.24
mg	-0.05	0.11	0.78	-0.11	0.52	0.08	0.57	0.04	-0.11	0.46	-0.24
mg/al	0.09	0.11	0.68	-0.01	0.45	0.06	0.63	-0.06	-0.02	0.34	-0.16
mn	-0.06	0.10	0.60	-0.02	0.48	0.09	0.52	0.00	-0.05	0.46	-0.21
mn/al	-0.04	0.06	0.41	0.09	0.36	0.06	0.48	-0.10	0.06	0.27	-0.11
mo	-0.05	-0.04	0.07	0.05	0.05	0.05	0.10	0.11	-0.09	-0.02	0.09
na	0.05	0.22	0.67	-0.08	0.30	0.10	0.60	-0.03	-0.06	0.35	-0.11
na/al	0.31	-0.14	-0.35	0.21	-0.20	-0.05	-0.20	-0.09	0.17	-0.32	0.19
nb	0.03	-0.01	0.06	0.11	0.13	-0.03	0.04	0.01	0.12	0.08	0.00
ni	-0.01	0.02	0.08	0.06	0.13	0.06	0.01	0.21	-0.22	0.03	0.01
p	-0.04	0.14	0.81	-0.10	0.54	0.08	0.66	0.01	-0.11	0.49	-0.23
pb	-0.03	0.04	0.15	0.12	0.12	0.18	0.17	0.01	0.00	0.19	-0.02
rb	-0.03	0.28	0.46	-0.10	0.31	0.06	0.41	0.03	-0.14	0.48	-0.13
re	-0.05	-0.12	-0.18	0.19	-0.15	0.18	-0.08	0.16	0.09	-0.11	0.11
s	0.07	-0.28	-0.28	0.42	0.01	-0.02	-0.46	0.20	0.01	-0.07	0.12
sb	-0.01	-0.08	-0.06	0.29	0.00	0.04	-0.11	0.01	0.06	-0.04	0.10
sc	-0.05	0.14	0.69	-0.09	0.42	0.13	0.50	0.05	-0.11	0.42	-0.18
se	0.02	-0.06	-0.03	0.44	-0.07	0.03	0.01	0.04	0.25	-0.22	0.60
sn	0.04	-0.11	-0.09	0.28	0.05	-0.02	-0.20	0.02	0.10	-0.07	0.11
sr	0.07	0.08	0.47	-0.01	0.07	0.01	0.19	0.01	-0.08	0.21	0.04

Tabla A.6: Matriz de correlación. Parte 3 de 7

	b	ba	be	bi	ca	cd	ce	co	cr	cs	cu
ta	-0.01	0.00	0.02	0.01	0.00	-0.01	0.04	-0.02	-0.02	0.05	0.02
te	0.15	-0.13	-0.17	0.41	0.03	-0.01	-0.27	0.03	0.10	-0.08	0.20
th	-0.09	0.21	0.63	-0.08	0.27	0.04	0.79	-0.15	0.00	0.27	-0.15
ti	-0.02	0.18	0.31	-0.07	0.30	-0.01	0.24	0.00	-0.06	0.28	-0.13
tl	0.00	-0.06	-0.19	0.19	-0.11	0.10	-0.21	0.28	0.05	-0.03	0.16
u	-0.06	0.06	0.17	0.27	-0.04	0.27	0.33	0.04	0.07	0.18	0.05
v	-0.03	0.13	0.72	-0.11	0.49	0.12	0.50	0.06	-0.10	0.48	-0.21
w	0.01	-0.01	-0.03	0.07	0.00	0.01	-0.05	0.03	0.01	0.00	0.05
y	-0.07	0.13	0.79	-0.06	0.38	0.10	0.63	-0.03	-0.08	0.42	-0.10
zn	-0.07	0.07	0.41	0.18	0.24	0.24	0.45	0.04	0.05	0.28	-0.02
zr	0.00	0.00	0.08	0.00	0.00	0.00	0.04	0.01	-0.01	0.02	0.01

Tabla A.7: Matriz de correlación. Parte 4 de 7

	cu/s	fe	ga	ge	hf	hg	in	k	k/al	k/mg	kxal
(al+k)/(na+ca+mg)	-0.11	-0.23	-0.53	-0.25	-0.12	0.02	0.04	-0.10	0.32	0.59	-0.35
(al+k+na)/(ca+mg)	-0.08	-0.24	-0.51	-0.24	-0.12	0.01	0.06	-0.14	0.27	0.55	-0.34
(ca+na)/(k+al)	-0.11	0.22	-0.05	0.15	-0.01	0.02	-0.05	-0.13	0.15	0.00	-0.07
(k+al+s)/(fe+s)	0.53	-0.49	0.28	-0.04	0.08	-0.11	-0.08	0.38	-0.36	-0.22	0.36
(kxna)/al	0.16	-0.23	-0.08	-0.07	-0.01	0.00	0.06	0.37	0.34	0.02	0.09
3*al/(na+k)	0.10	0.27	0.78	0.22	0.15	-0.08	-0.04	0.16	-0.84	-0.59	0.60
ag	0.04	0.16	-0.01	0.22	0.00	0.41	0.46	-0.08	-0.02	-0.11	-0.08
al	0.15	0.25	0.92	0.26	0.20	-0.10	-0.04	0.56	-0.76	-0.59	0.89
al/(na+ca+k)	-0.13	-0.18	-0.42	-0.21	-0.10	0.03	0.03	0.02	0.39	0.49	-0.27
al/mg	-0.23	-0.17	-0.66	-0.24	-0.13	0.04	-0.04	-0.23	0.50	0.89	-0.46
as	-0.13	0.24	-0.11	0.48	-0.03	0.38	0.11	-0.13	0.07	0.06	-0.10
au	0.05	-0.01	-0.02	0.02	-0.03	-0.02	0.05	-0.04	-0.02	-0.04	-0.04
b	-0.01	0.03	-0.04	0.00	0.06	0.01	-0.01	-0.01	0.08	0.01	-0.04
ba	0.19	-0.18	0.18	-0.04	0.03	-0.07	-0.06	0.38	-0.09	-0.15	0.27
be	0.19	0.17	0.80	0.28	0.20	-0.06	-0.01	0.43	-0.66	-0.65	0.70
bi	-0.07	0.31	-0.10	0.24	-0.03	0.24	0.42	-0.17	0.04	-0.04	-0.16
ca	-0.01	0.22	0.47	0.19	0.09	-0.06	-0.04	0.31	-0.22	-0.28	0.46
cd	0.04	0.03	0.09	0.01	0.01	0.09	0.25	0.04	-0.12	-0.10	0.08
ce	0.33	-0.14	0.60	0.16	0.12	-0.09	0.11	0.37	-0.53	-0.62	0.46
co	-0.28	0.23	0.02	0.04	0.01	0.05	0.15	0.02	-0.01	0.09	0.07
cr	0.11	-0.03	-0.09	0.04	-0.04	-0.01	0.07	-0.20	0.04	-0.05	-0.17
cs	-0.04	0.21	0.48	0.13	0.13	-0.01	0.10	0.40	-0.39	-0.38	0.51
cu	0.44	-0.11	-0.23	0.07	-0.02	0.12	0.16	-0.15	0.12	0.03	-0.23
cu/s	1.00	-0.45	0.26	0.00	0.05	-0.08	-0.13	0.14	-0.27	-0.33	0.17
fe	-0.45	1.00	0.24	0.48	0.05	0.14	0.08	0.00	-0.15	-0.11	0.22
ga	0.26	0.24	1.00	0.32	0.21	-0.08	-0.03	0.55	-0.71	-0.66	0.86
ge	0.00	0.48	0.32	1.00	0.15	0.10	0.04	0.11	-0.24	-0.24	0.27
hf	0.05	0.05	0.21	0.15	1.00	-0.02	-0.02	0.13	-0.14	-0.13	0.20
hg	-0.08	0.14	-0.08	0.10	-0.02	1.00	0.25	-0.06	0.08	0.05	-0.08
in	-0.13	0.08	-0.03	0.04	-0.02	0.25	1.00	0.00	-0.02	-0.07	-0.05
k	0.14	0.00	0.55	0.11	0.13	-0.06	0.00	1.00	-0.14	-0.18	0.81

Tabla A.7: Matriz de correlación. Parte 4 de 7

	cu/s	fe	ga	ge	hf	hg	in	k	k/al	k/mg	kxal
k/al	-0.27	-0.15	-0.71	-0.24	-0.14	0.08	-0.02	-0.14	1.00	0.76	-0.54
k/mg	-0.33	-0.11	-0.66	-0.24	-0.13	0.05	-0.07	-0.18	0.76	1.00	-0.48
kxal	0.17	0.22	0.86	0.27	0.20	-0.08	-0.05	0.81	-0.54	-0.48	1.00
la	0.33	-0.18	0.54	0.13	0.11	-0.09	0.12	0.35	-0.49	-0.58	0.41
li	0.18	0.27	0.93	0.31	0.19	-0.10	-0.06	0.50	-0.67	-0.62	0.83
mg	0.18	0.28	0.96	0.33	0.20	-0.10	-0.07	0.51	-0.66	-0.62	0.85
mg/al	0.26	0.18	0.83	0.28	0.16	-0.09	-0.05	0.36	-0.60	-0.71	0.62
mn	0.10	0.22	0.77	0.25	0.14	-0.08	0.04	0.37	-0.53	-0.54	0.61
mn/al	0.15	0.12	0.53	0.20	0.08	-0.05	0.05	0.14	-0.37	-0.51	0.30
mo	-0.02	-0.02	0.00	0.11	0.03	0.11	0.12	0.05	-0.04	-0.01	0.06
na	0.37	0.04	0.75	0.21	0.16	-0.09	0.00	0.49	-0.66	-0.65	0.72
na/al	0.02	-0.09	-0.40	-0.02	-0.09	0.05	0.03	-0.40	0.34	0.09	-0.43
nb	0.12	0.25	0.22	0.41	0.20	0.02	-0.04	0.20	-0.03	-0.08	0.20
ni	0.03	0.17	0.16	0.06	0.15	0.01	0.04	0.09	-0.11	-0.09	0.15
p	0.15	0.25	0.85	0.31	0.18	-0.07	-0.03	0.46	-0.64	-0.61	0.74
pb	-0.02	0.04	0.19	0.04	0.05	0.33	0.19	0.10	-0.12	-0.18	0.11
rb	0.24	0.05	0.66	0.19	0.16	-0.05	0.01	0.89	-0.26	-0.29	0.82
re	-0.07	-0.04	-0.17	-0.06	-0.03	0.18	0.36	-0.09	0.07	0.11	-0.12
s	-0.62	0.74	-0.32	0.31	-0.05	0.22	0.18	-0.26	0.27	0.28	-0.27
sb	-0.10	0.24	-0.06	0.37	-0.02	0.41	0.11	-0.11	0.04	0.01	-0.07
sc	0.26	0.21	0.88	0.30	0.22	-0.08	-0.02	0.59	-0.59	-0.54	0.85
se	0.23	0.12	-0.10	0.30	0.02	0.15	0.21	-0.13	-0.05	-0.11	-0.12
sn	-0.15	0.47	-0.07	0.31	-0.01	0.10	0.01	-0.12	0.07	0.05	-0.06
sr	-0.05	0.09	0.15	0.10	0.09	0.04	0.08	0.10	-0.26	-0.18	0.17
ta	0.05	-0.03	0.01	0.08	0.23	0.25	-0.01	0.01	0.01	0.00	0.00
te	-0.18	0.51	-0.15	0.40	-0.03	0.17	0.09	-0.21	0.14	0.08	-0.16
th	0.30	-0.09	0.60	0.12	0.10	-0.13	-0.01	0.32	-0.50	-0.62	0.43
ti	0.29	0.11	0.55	0.23	0.11	-0.04	-0.10	0.64	-0.19	-0.26	0.66
tl	-0.07	0.02	-0.13	0.00	0.00	0.14	0.26	0.11	0.16	0.24	0.01
u	0.01	-0.04	0.09	0.01	0.03	0.14	0.42	0.08	-0.15	-0.22	0.06
v	0.22	0.30	0.92	0.33	0.22	-0.09	-0.06	0.58	-0.63	-0.57	0.88
w	-0.08	0.06	-0.07	0.04	-0.02	0.19	0.10	-0.01	0.08	0.06	-0.05
y	0.24	0.13	0.75	0.27	0.34	-0.09	-0.02	0.41	-0.57	-0.60	0.64
zn	0.10	0.09	0.48	0.17	0.06	0.23	0.39	0.23	-0.37	-0.42	0.34
zr	-0.01	0.02	0.03	0.06	0.06	0.00	0.00	0.02	-0.03	-0.03	0.05

Tabla A.8: Matriz de correlación. Parte 5 de 7

	la	li	mg	mg/al	mn	mn/al	mo	na	na/al	nb	ni
(al+k)/(na+ca+mg)	-0.33	-0.54	-0.57	-0.66	-0.50	-0.52	0.00	-0.45	-0.07	-0.13	-0.10
(al+k+na)/(ca+mg)	-0.30	-0.51	-0.54	-0.62	-0.47	-0.48	0.03	-0.36	0.03	-0.12	-0.09
(ca+na)/(k+al)	-0.14	-0.01	0.00	0.07	0.04	0.16	-0.02	-0.09	0.22	0.08	0.06
(k+al+s)/(fe+s)	0.33	0.22	0.21	0.13	0.11	0.00	0.04	0.36	-0.25	-0.06	-0.05
(kxna)/al	0.13	-0.12	-0.11	0.05	-0.10	-0.02	0.04	0.37	0.61	0.09	0.01
3*al/(na+k)	0.37	0.77	0.78	0.59	0.58	0.32	0.05	0.52	-0.49	0.02	0.12
ag	0.07	-0.06	-0.06	0.00	0.07	0.16	0.00	-0.01	0.10	0.09	0.06

Tabla A.8: Matriz de correlación. Parte 5 de 7

	la	li	mg	mg/al	mn	mn/al	mo	na	na/al	nb	ni
al	0.47	0.90	0.91	0.68	0.68	0.35	0.06	0.73	-0.51	0.10	0.15
al/(na+ca+k)	-0.24	-0.45	-0.45	-0.49	-0.42	-0.42	-0.01	-0.45	-0.12	-0.10	-0.07
al/mg	-0.55	-0.64	-0.66	-0.77	-0.58	-0.57	-0.01	-0.57	0.06	-0.10	-0.10
as	-0.15	-0.11	-0.11	-0.10	-0.09	-0.03	0.01	-0.12	0.16	0.08	-0.02
au	0.06	-0.04	-0.04	-0.02	-0.05	-0.02	-0.02	0.01	0.02	0.04	0.03
b	-0.10	-0.03	-0.05	0.09	-0.06	-0.04	-0.05	0.05	0.31	0.03	-0.01
ba	0.23	0.13	0.11	0.11	0.10	0.06	-0.04	0.22	-0.14	-0.01	0.02
be	0.58	0.78	0.78	0.68	0.60	0.41	0.07	0.67	-0.35	0.06	0.08
bi	0.00	-0.12	-0.11	-0.01	-0.02	0.09	0.05	-0.08	0.21	0.11	0.06
ca	0.18	0.49	0.52	0.45	0.48	0.36	0.05	0.30	-0.20	0.13	0.13
cd	0.10	0.09	0.08	0.06	0.09	0.06	0.05	0.10	-0.05	-0.03	0.06
ce	0.99	0.58	0.57	0.63	0.52	0.48	0.10	0.60	-0.20	0.04	0.01
co	-0.12	0.04	0.04	-0.06	0.00	-0.10	0.11	-0.03	-0.09	0.01	0.21
cr	-0.02	-0.10	-0.11	-0.02	-0.05	0.06	-0.09	-0.06	0.17	0.12	-0.22
cs	0.33	0.51	0.46	0.34	0.46	0.27	-0.02	0.35	-0.32	0.08	0.03
cu	-0.09	-0.24	-0.24	-0.16	-0.21	-0.11	0.09	-0.11	0.19	0.00	0.01
cu/s	0.33	0.18	0.18	0.26	0.10	0.15	-0.02	0.37	0.02	0.12	0.03
fe	-0.18	0.27	0.28	0.18	0.22	0.12	-0.02	0.04	-0.09	0.25	0.17
ga	0.54	0.93	0.96	0.83	0.77	0.53	0.00	0.75	-0.40	0.22	0.16
ge	0.13	0.31	0.33	0.28	0.25	0.20	0.11	0.21	-0.02	0.41	0.06
hf	0.11	0.19	0.20	0.16	0.14	0.08	0.03	0.16	-0.09	0.20	0.15
hg	-0.09	-0.10	-0.10	-0.09	-0.08	-0.05	0.11	-0.09	0.05	0.02	0.01
in	0.12	-0.06	-0.07	-0.05	0.04	0.05	0.12	0.00	0.03	-0.04	0.04
k	0.35	0.50	0.51	0.36	0.37	0.14	0.05	0.49	-0.40	0.20	0.09
k/al	-0.49	-0.67	-0.66	-0.60	-0.53	-0.37	-0.04	-0.66	0.34	-0.03	-0.11
k/mg	-0.58	-0.62	-0.62	-0.71	-0.54	-0.51	-0.01	-0.65	0.09	-0.08	-0.09
kxal	0.41	0.83	0.85	0.62	0.61	0.30	0.06	0.72	-0.43	0.20	0.15
la	1.00	0.52	0.51	0.57	0.48	0.46	0.10	0.55	-0.17	0.02	0.00
li	0.52	1.00	0.96	0.82	0.78	0.51	0.06	0.69	-0.38	0.16	0.14
mg	0.51	0.96	1.00	0.86	0.78	0.52	0.06	0.71	-0.38	0.19	0.17
mg/al	0.57	0.82	0.86	1.00	0.74	0.68	0.01	0.67	-0.08	0.20	0.14
mn	0.48	0.78	0.78	0.74	1.00	0.84	0.04	0.52	-0.30	0.15	0.12
mn/al	0.46	0.51	0.52	0.68	0.84	1.00	-0.02	0.36	-0.06	0.13	0.12
mo	0.10	0.06	0.06	0.01	0.04	-0.02	1.00	0.06	-0.02	-0.04	0.00
na	0.55	0.69	0.71	0.67	0.52	0.36	0.06	1.00	0.01	0.12	0.14
na/al	-0.17	-0.38	-0.38	-0.08	-0.30	-0.06	-0.02	0.01	1.00	-0.04	-0.03
nb	0.02	0.16	0.19	0.20	0.15	0.13	-0.04	0.12	-0.04	1.00	0.04
ni	0.00	0.14	0.17	0.14	0.12	0.12	0.00	0.14	-0.03	0.04	1.00
p	0.60	0.87	0.87	0.76	0.72	0.52	0.08	0.66	-0.37	0.14	0.10
pb	0.17	0.16	0.14	0.17	0.34	0.37	0.00	0.10	-0.07	-0.01	0.01
rb	0.38	0.58	0.60	0.46	0.45	0.23	0.01	0.54	-0.37	0.33	0.08
re	-0.07	-0.15	-0.16	-0.19	-0.14	-0.15	0.28	-0.09	0.08	-0.05	-0.02
s	-0.46	-0.28	-0.27	-0.29	-0.20	-0.15	-0.01	-0.40	0.14	0.09	0.04
sb	-0.11	-0.07	-0.07	-0.05	-0.03	0.04	0.01	-0.09	0.15	0.07	0.00
sc	0.44	0.82	0.84	0.68	0.60	0.34	0.04	0.73	-0.35	0.27	0.18
se	0.02	-0.11	-0.11	-0.03	-0.10	-0.01	0.19	0.03	0.19	0.08	0.07

Tabla A.8: Matriz de correlación. Parte 5 de 7

	la	li	mg	mg/al	mn	mn/al	mo	na	na/al	nb	ni
sn	-0.20	-0.06	-0.05	-0.05	-0.05	-0.01	-0.01	-0.11	0.14	0.21	0.07
sr	0.17	0.10	0.10	0.04	-0.01	-0.09	0.07	0.15	-0.18	-0.06	-0.02
ta	0.04	0.01	0.01	0.01	0.03	0.02	0.06	-0.02	-0.02	0.07	-0.01
te	-0.27	-0.14	-0.14	-0.06	-0.10	-0.01	-0.06	-0.18	0.27	0.16	0.03
th	0.78	0.55	0.58	0.66	0.54	0.53	0.04	0.58	-0.16	0.06	0.03
ti	0.21	0.48	0.51	0.43	0.35	0.21	-0.01	0.50	-0.14	0.54	0.11
tl	-0.20	-0.13	-0.14	-0.24	-0.14	-0.21	0.12	-0.10	-0.04	0.03	0.04
u	0.34	0.05	0.04	0.07	0.07	0.10	0.12	0.17	0.00	-0.06	-0.02
v	0.44	0.90	0.92	0.75	0.69	0.42	0.03	0.75	-0.37	0.26	0.17
w	-0.05	-0.08	-0.08	-0.07	-0.06	-0.04	0.01	-0.08	0.00	0.00	0.01
y	0.58	0.76	0.76	0.71	0.60	0.45	0.08	0.64	-0.27	0.12	0.13
zn	0.43	0.45	0.44	0.46	0.56	0.52	0.00	0.37	-0.15	0.03	0.09
zr	0.03	0.01	0.03	0.01	0.00	-0.01	0.00	0.04	-0.02	0.00	0.00

Tabla A.9: Matriz de correlación. Parte 6 de 7

	p	pb	rb	re	s	sb	sc	se	sn	sr	ta
(al+k)/(na+ca+mg)	-0.55	-0.12	-0.18	0.21	0.08	-0.04	-0.43	-0.01	-0.06	-0.13	-0.01
(al+k+na)/(ca+mg)	-0.53	-0.11	-0.20	0.24	0.05	-0.04	-0.41	0.03	-0.06	-0.16	-0.02
(ca+na)/(k+al)	0.05	0.00	-0.12	-0.11	0.28	0.18	-0.05	0.05	0.26	0.00	-0.01
(k+al+s)/(fe+s)	0.21	0.04	0.34	-0.05	-0.49	-0.10	0.28	-0.02	-0.19	0.13	0.04
(kxna)/al	-0.10	-0.01	0.28	0.03	-0.15	-0.04	0.01	0.09	-0.05	-0.11	-0.01
3*al/(na+k)	0.72	0.09	0.24	-0.09	-0.23	-0.07	0.62	-0.11	-0.09	0.28	0.00
ag	-0.03	0.39	-0.02	0.15	0.26	0.36	-0.05	0.43	0.17	-0.01	0.01
al	0.83	0.12	0.60	-0.12	-0.30	-0.10	0.82	-0.13	-0.10	0.26	0.00
al/(na+ca+k)	-0.49	-0.10	-0.08	0.15	0.07	-0.04	-0.34	-0.04	-0.06	-0.12	-0.01
al/mg	-0.62	-0.17	-0.30	0.17	0.23	0.00	-0.53	-0.04	0.03	-0.12	-0.01
as	-0.09	0.08	-0.12	0.04	0.37	0.79	-0.10	0.15	0.22	0.00	0.05
au	-0.05	-0.01	-0.06	-0.03	-0.01	-0.02	-0.05	0.35	-0.01	0.03	-0.01
b	-0.04	-0.03	-0.03	-0.05	0.07	-0.01	-0.05	0.02	0.04	0.07	-0.01
ba	0.14	0.04	0.28	-0.12	-0.28	-0.08	0.14	-0.06	-0.11	0.08	0.00
be	0.81	0.15	0.46	-0.18	-0.28	-0.06	0.69	-0.03	-0.09	0.47	0.02
bi	-0.10	0.12	-0.10	0.19	0.42	0.29	-0.09	0.44	0.28	-0.01	0.01
ca	0.54	0.12	0.31	-0.15	0.01	0.00	0.42	-0.07	0.05	0.07	0.00
cd	0.08	0.18	0.06	0.18	-0.02	0.04	0.13	0.03	-0.02	0.01	-0.01
ce	0.66	0.17	0.41	-0.08	-0.46	-0.11	0.50	0.01	-0.20	0.19	0.04
co	0.01	0.01	0.03	0.16	0.20	0.01	0.05	0.04	0.02	0.01	-0.02
cr	-0.11	0.00	-0.14	0.09	0.01	0.06	-0.11	0.25	0.10	-0.08	-0.02
cs	0.49	0.19	0.48	-0.11	-0.07	-0.04	0.42	-0.22	-0.07	0.21	0.05
cu	-0.23	-0.02	-0.13	0.11	0.12	0.10	-0.18	0.60	0.11	0.04	0.02
cu/s	0.15	-0.02	0.24	-0.07	-0.62	-0.10	0.26	0.23	-0.15	-0.05	0.05
fe	0.25	0.04	0.05	-0.04	0.74	0.24	0.21	0.12	0.47	0.09	-0.03
ga	0.85	0.19	0.66	-0.17	-0.32	-0.06	0.88	-0.10	-0.07	0.15	0.01
ge	0.31	0.04	0.19	-0.06	0.31	0.37	0.30	0.30	0.31	0.10	0.08
hf	0.18	0.05	0.16	-0.03	-0.05	-0.02	0.22	0.02	-0.01	0.09	0.23
hg	-0.07	0.33	-0.05	0.18	0.22	0.41	-0.08	0.15	0.10	0.04	0.25

Tabla A.9: Matriz de correlación. Parte 6 de 7

	p	pb	rb	re	s	sb	sc	se	sn	sr	ta
in	-0.03	0.19	0.01	0.36	0.18	0.11	-0.02	0.21	0.01	0.08	-0.01
k	0.46	0.10	0.89	-0.09	-0.26	-0.11	0.59	-0.13	-0.12	0.10	0.01
k/al	-0.64	-0.12	-0.26	0.07	0.27	0.04	-0.59	-0.05	0.07	-0.26	0.01
k/mg	-0.61	-0.18	-0.29	0.11	0.28	0.01	-0.54	-0.11	0.05	-0.18	0.00
kxal	0.74	0.11	0.82	-0.12	-0.27	-0.07	0.85	-0.12	-0.06	0.17	0.00
la	0.60	0.17	0.38	-0.07	-0.46	-0.11	0.44	0.02	-0.20	0.17	0.04
li	0.87	0.16	0.58	-0.15	-0.28	-0.07	0.82	-0.11	-0.06	0.10	0.01
mg	0.87	0.14	0.60	-0.16	-0.27	-0.07	0.84	-0.11	-0.05	0.10	0.01
mg/al	0.76	0.17	0.46	-0.19	-0.29	-0.05	0.68	-0.03	-0.05	0.04	0.01
mn	0.72	0.34	0.45	-0.14	-0.20	-0.03	0.60	-0.10	-0.05	-0.01	0.03
mn/al	0.52	0.37	0.23	-0.15	-0.15	0.04	0.34	-0.01	-0.01	-0.09	0.02
mo	0.08	0.00	0.01	0.28	-0.01	0.01	0.04	0.19	-0.01	0.07	0.06
na	0.66	0.10	0.54	-0.09	-0.40	-0.09	0.73	0.03	-0.11	0.15	-0.02
na/al	-0.37	-0.07	-0.37	0.08	0.14	0.15	-0.35	0.19	0.14	-0.18	-0.02
nb	0.14	-0.01	0.33	-0.05	0.09	0.07	0.27	0.08	0.21	-0.06	0.07
ni	0.10	0.01	0.08	-0.02	0.04	0.00	0.18	0.07	0.07	-0.02	-0.01
p	1.00	0.18	0.52	-0.17	-0.24	-0.06	0.73	-0.09	-0.06	0.31	0.01
pb	0.18	1.00	0.13	-0.03	0.00	0.17	0.11	0.03	-0.02	-0.04	0.00
rb	0.52	0.13	1.00	-0.05	-0.28	-0.09	0.71	-0.11	-0.09	0.04	0.01
re	-0.17	-0.03	-0.05	1.00	0.06	0.03	-0.13	0.15	0.00	-0.05	-0.01
s	-0.24	0.00	-0.28	0.06	1.00	0.32	-0.28	0.22	0.48	0.03	-0.03
sb	-0.06	0.17	-0.09	0.03	0.32	1.00	-0.06	0.15	0.27	-0.03	0.02
sc	0.73	0.11	0.71	-0.13	-0.28	-0.06	1.00	-0.07	-0.05	0.13	-0.01
se	-0.09	0.03	-0.11	0.15	0.22	0.15	-0.07	1.00	0.20	0.09	0.03
sn	-0.06	-0.02	-0.09	0.00	0.48	0.27	-0.05	0.20	1.00	-0.04	0.00
sr	0.31	-0.04	0.04	-0.05	0.03	-0.03	0.13	0.09	-0.04	1.00	0.00
ta	0.01	0.00	0.01	-0.01	-0.03	0.02	-0.01	0.03	0.00	0.00	1.00
te	-0.14	0.00	-0.17	-0.01	0.61	0.39	-0.14	0.26	0.60	-0.02	-0.01
th	0.59	0.14	0.30	-0.18	-0.42	-0.11	0.47	-0.03	-0.17	0.15	0.01
ti	0.41	0.03	0.77	-0.09	-0.22	-0.03	0.66	-0.06	0.02	-0.07	0.00
tl	-0.17	-0.01	0.16	0.59	0.11	0.01	-0.05	0.15	0.03	-0.04	0.01
u	0.10	0.14	0.08	0.23	-0.06	0.06	0.07	0.13	-0.07	0.20	0.00
v	0.81	0.12	0.69	-0.14	-0.26	-0.06	0.93	-0.10	-0.01	0.11	0.00
w	-0.04	0.10	-0.03	0.02	0.12	0.08	-0.08	0.06	0.05	0.09	-0.01
y	0.76	0.13	0.48	-0.16	-0.28	-0.05	0.70	0.01	-0.07	0.23	0.00
zn	0.45	0.64	0.29	-0.04	-0.11	0.15	0.40	0.10	-0.05	-0.01	0.01
zr	0.10	0.00	0.03	-0.01	0.01	0.00	0.05	0.02	0.00	0.16	0.00

Tabla A.10: Matriz de correlación. Parte 7 de 7

	te	th	ti	tl	u	v	w	y	zn	zr
(al+k)/(na+ca+mg)	-0.06	-0.46	-0.24	0.27	-0.05	-0.50	0.00	-0.53	-0.30	-0.02
(al+k+na)/(ca+mg)	-0.06	-0.43	-0.22	0.26	-0.03	-0.47	-0.02	-0.50	-0.28	-0.02
(ca+na)/(k+al)	0.32	-0.07	0.01	-0.12	-0.10	-0.01	0.04	-0.01	-0.02	-0.01
(k+al+s)/(fe+s)	-0.23	0.26	0.14	0.03	0.06	0.21	-0.05	0.23	0.13	0.02
(kxna)/al	-0.03	0.11	0.31	0.07	0.11	-0.03	0.00	-0.02	0.00	-0.01

Tabla A.10: Matriz de correlación. Parte 7 de 7

	te	th	ti	tl	u	v	w	y	zn	zr
3*al/(na+k)	-0.15	0.42	0.14	-0.14	0.07	0.69	-0.07	0.57	0.32	0.04
ag	0.28	-0.01	-0.05	0.16	0.32	-0.06	0.15	0.00	0.45	0.00
al	-0.19	0.50	0.42	-0.07	0.09	0.88	-0.07	0.70	0.39	0.06
al/(na+ca+k)	-0.08	-0.32	-0.17	0.21	-0.03	-0.41	0.00	-0.41	-0.25	-0.01
al/mg	0.06	-0.64	-0.27	0.28	-0.17	-0.58	0.04	-0.63	-0.41	-0.03
as	0.39	-0.16	-0.05	0.05	0.04	-0.09	0.08	-0.10	0.05	0.00
au	-0.01	0.06	-0.04	-0.02	0.14	-0.05	0.00	-0.04	-0.02	0.00
b	0.15	-0.09	-0.02	0.00	-0.06	-0.03	0.01	-0.07	-0.07	0.00
ba	-0.13	0.21	0.18	-0.06	0.06	0.13	-0.01	0.13	0.07	0.00
be	-0.17	0.63	0.31	-0.19	0.17	0.72	-0.03	0.79	0.41	0.08
bi	0.41	-0.08	-0.07	0.19	0.27	-0.11	0.07	-0.06	0.18	0.00
ca	0.03	0.27	0.30	-0.11	-0.04	0.49	0.00	0.38	0.24	0.00
cd	-0.01	0.04	-0.01	0.10	0.27	0.12	0.01	0.10	0.24	0.00
ce	-0.27	0.79	0.24	-0.21	0.33	0.50	-0.05	0.63	0.45	0.04
co	0.03	-0.15	0.00	0.28	0.04	0.06	0.03	-0.03	0.04	0.01
cr	0.10	0.00	-0.06	0.05	0.07	-0.10	0.01	-0.08	0.05	-0.01
cs	-0.08	0.27	0.28	-0.03	0.18	0.48	0.00	0.42	0.28	0.02
cu	0.20	-0.15	-0.13	0.16	0.05	-0.21	0.05	-0.10	-0.02	0.01
cu/s	-0.18	0.30	0.29	-0.07	0.01	0.22	-0.08	0.24	0.10	-0.01
fe	0.51	-0.09	0.11	0.02	-0.04	0.30	0.06	0.13	0.09	0.02
ga	-0.15	0.60	0.55	-0.13	0.09	0.92	-0.07	0.75	0.48	0.03
ge	0.40	0.12	0.23	0.00	0.01	0.33	0.04	0.27	0.17	0.06
hf	-0.03	0.10	0.11	0.00	0.03	0.22	-0.02	0.34	0.06	0.06
hg	0.17	-0.13	-0.04	0.14	0.14	-0.09	0.19	-0.09	0.23	0.00
in	0.09	-0.01	-0.10	0.26	0.42	-0.06	0.10	-0.02	0.39	0.00
k	-0.21	0.32	0.64	0.11	0.08	0.58	-0.01	0.41	0.23	0.02
k/al	0.14	-0.50	-0.19	0.16	-0.15	-0.63	0.08	-0.57	-0.37	-0.03
k/mg	0.08	-0.62	-0.26	0.24	-0.22	-0.57	0.06	-0.60	-0.42	-0.03
kxal	-0.16	0.43	0.66	0.01	0.06	0.88	-0.05	0.64	0.34	0.05
la	-0.27	0.78	0.21	-0.20	0.34	0.44	-0.05	0.58	0.43	0.03
li	-0.14	0.55	0.48	-0.13	0.05	0.90	-0.08	0.76	0.45	0.01
mg	-0.14	0.58	0.51	-0.14	0.04	0.92	-0.08	0.76	0.44	0.03
mg/al	-0.06	0.66	0.43	-0.24	0.07	0.75	-0.07	0.71	0.46	0.01
mn	-0.10	0.54	0.35	-0.14	0.07	0.69	-0.06	0.60	0.56	0.00
mn/al	-0.01	0.53	0.21	-0.21	0.10	0.42	-0.04	0.45	0.52	-0.01
mo	-0.06	0.04	-0.01	0.12	0.12	0.03	0.01	0.08	0.00	0.00
na	-0.18	0.58	0.50	-0.10	0.17	0.75	-0.08	0.64	0.37	0.04
na/al	0.27	-0.16	-0.14	-0.04	0.00	-0.37	0.00	-0.27	-0.15	-0.02
nb	0.16	0.06	0.54	0.03	-0.06	0.26	0.00	0.12	0.03	0.00
ni	0.03	0.03	0.11	0.04	-0.02	0.17	0.01	0.13	0.09	0.00
p	-0.14	0.59	0.41	-0.17	0.10	0.81	-0.04	0.76	0.45	0.10
pb	0.00	0.14	0.03	-0.01	0.14	0.12	0.10	0.13	0.64	0.00
rb	-0.17	0.30	0.77	0.16	0.08	0.69	-0.03	0.48	0.29	0.03
re	-0.01	-0.18	-0.09	0.59	0.23	-0.14	0.02	-0.16	-0.04	-0.01
s	0.61	-0.42	-0.22	0.11	-0.06	-0.26	0.12	-0.28	-0.11	0.01
sb	0.39	-0.11	-0.03	0.01	0.06	-0.06	0.08	-0.05	0.15	0.00

Tabla A.10: Matriz de correlación. Parte 7 de 7

	te	th	ti	tl	u	v	w	y	zn	zr
sc	-0.14	0.47	0.66	-0.05	0.07	0.93	-0.08	0.70	0.40	0.05
se	0.26	-0.03	-0.06	0.15	0.13	-0.10	0.06	0.01	0.10	0.02
sn	0.60	-0.17	0.02	0.03	-0.07	-0.01	0.05	-0.07	-0.05	0.00
sr	-0.02	0.15	-0.07	-0.04	0.20	0.11	0.09	0.23	-0.01	0.16
ta	-0.01	0.01	0.00	0.01	0.00	0.00	-0.01	0.00	0.01	0.00
te	1.00	-0.26	-0.07	0.05	-0.07	-0.12	0.07	-0.14	-0.03	0.00
th	-0.26	1.00	0.25	-0.33	0.27	0.48	-0.05	0.62	0.39	0.00
ti	-0.07	0.25	1.00	0.04	-0.04	0.65	-0.05	0.37	0.16	-0.01
tl	0.05	-0.33	0.04	1.00	0.11	-0.07	0.12	-0.18	-0.06	-0.01
u	-0.07	0.27	-0.04	0.11	1.00	0.04	0.04	0.17	0.25	0.01
v	-0.12	0.48	0.65	-0.07	0.04	1.00	-0.08	0.72	0.41	0.03
w	0.07	-0.05	-0.05	0.12	0.04	-0.08	1.00	-0.06	0.02	0.01
y	-0.14	0.62	0.37	-0.18	0.17	0.72	-0.06	1.00	0.42	0.02
zn	-0.03	0.39	0.16	-0.06	0.25	0.41	0.02	0.42	1.00	0.01
zr	0.00	0.00	-0.01	-0.01	0.01	0.03	0.01	0.02	0.01	1.00

Anexo B

Antecedentes

B.1. Clustering jerárquico

En la minería de datos, el clustering jerárquico es un método de agrupamiento que determina los clusters en base a una jerarquía establecida, las cuales son principalmente de dos tipos:

- *Agglomerativa*: también denominada «de abajo hacia arriba». Cada dato empieza contenido en su propio cluster y se empiezan a unir en clusters más grandes a medida que aumenta su jerarquía.
- *Divisiva*: también denominada «de arriba hacia abajo». Todos los datos se encuentran en un único cluster y se van separando recursivamente a medida que disminuye la jerarquía.

El cómo los clusters son combinados o separados depende del uso de una métrica, que define la distancia entre observaciones, además de un criterio de unión, el cual especifica qué tan disimiles son un cluster de otro. Entre las métricas más usadas se encuentran la distancia Euclideana, el cuadrado de la distancia Euclideana, la máxima distancia, entre otras. El criterio de unión dependerá del algoritmo a implementar.

Anexo C

Enfoque tradicional de definición de dominios geometalúrgicos

En este capítulo, a modo ilustrativo, se detalla la metodología tradicional de definición de dominios geometalúrgicos, basada en asociaciones geológicas generadas en función de los datos disponibles.

C.1. Dominios

Usando la base de datos final, se definen 4 dominios en base a las principales alteraciones presentes (Tabla 4.6):

- Dominio A: Alteración del tipo Clorita-Sericita con presencia de arcillas (CSA). Con variedad de litologías y mineralización.
- Dominio B: Alteración del tipo Clorita-Sericita-Cuarzo sin presencia de arcillas (CSC). Con variedad de litologías y mineralización.
- Dominio C: Alteración del tipo Cuarzo-Sericita sin presencia de arcillas (SCC). Con variedad de litologías y mineralización.
- Dominio D: Alteración del tipo Cuarzo-Sericita con presencia de arcillas (SCA). Con variedad de litologías y mineralización.

Se destaca que dentro de los dominios definidos no se menciona la alteración SSI (Sericita sin identificar) debido a que los datos asociados a esa alteración pertenecen a una codificación antigua, siendo que podrían pertenecer a cualquiera de otras alteraciones con presencia de sericita, por lo que se omitió del estudio.

C.2. Variables seleccionadas

A continuación se muestran las variables seleccionadas por cada dominio en base al criterio definido previamente (página 28).

C.2.1. Dominio A

Tabla C.1: Variables seleccionadas en el dominio A. Metodología tradicional.

Variable	Correlación
K/Al	0.54
Na/Al	0.43
Ni	0.36
(KxNa)/Al	0.27
Cu	0.18
Li	-0.45
Al	-0.48
3*Al/(Na+K)	-0.48

C.2.2. Dominio B

Tabla C.2: Variables seleccionadas en el dominio B. Metodología tradicional.

Variable	Correlación
K/Al	0.57
Na/Al	0.55
Bi	0.40
Ag	0.40
Se	0.38
Cu	0.37
Mg	-0.55
P	-0.56
Be	-0.57
Li	-0.59
3*Al/(Na+K)	-0.65
Al	-0.70

C.2.3. Dominio C

Tabla C.3: Variables seleccionadas en el dominio C. Metodología tradicional.

Variable	Correlación
K/Al	0.48
K/Mg	0.42
Al/Mg	0.33
S	0.20
Cu	0.19
3*Al/(Na+K)	-0.49
Mg/Al	-0.52
Al	-0.52

C.2.4. Dominio D

Tabla C.4: Variables seleccionadas en el dominio D. Metodología tradicional.

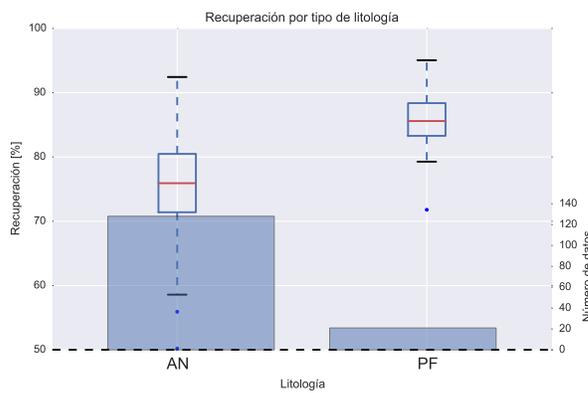
Variable	Correlación
K/Al	0.40
K/Mg	0.38
Al/Mg	0.33
Cu	0.26
S	0.24
3*Al/(Na+K)	-0.50
Mg	-0.51
Ga	-0.51
P	-0.52
V	-0.53
Al	-0.56
Be	-0.60

C.3. Estadísticas

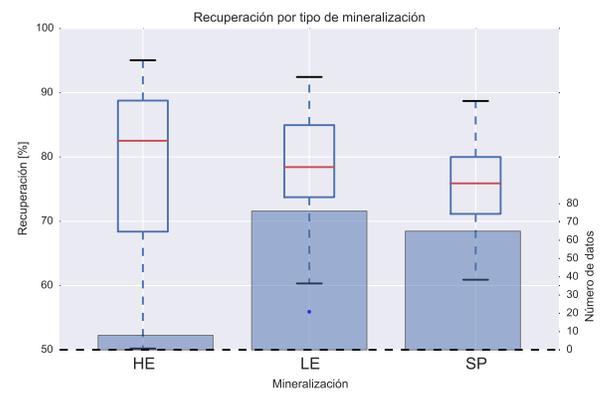
C.3.1. Dominio A

Tabla C.5: Estadísticas de recuperación. Dominio A de metodología clásica.

Número de datos	149
Media	77.06 %
Desviación estándar	8.16 %
CV	0.11
Mínimo	50.21 %
25 %	72.33 %
50 %	77.72 %
75 %	83.69 %
Máximo	95.05
RIC	11.35 %



(a)



(b)

Figura C.1: Boxplots de recuperación por tipo de variable geológica en el dominio A. Metodología original. (a): Recuperación por litologías. (b): Recuperación por mineralizaciones.

C.3.2. Dominio B

Tabla C.6: Estadísticas de recuperación. Dominio B de metodología clásica.

Número de datos	256
Media	77.78 %
Desviación estándar	8.24 %
CV	0.11
Mínimo	40 %
25 %	72.27 %
50 %	78.51 %
75 %	83.65 %
Máximo	94.33 %
RIC	11.38 %

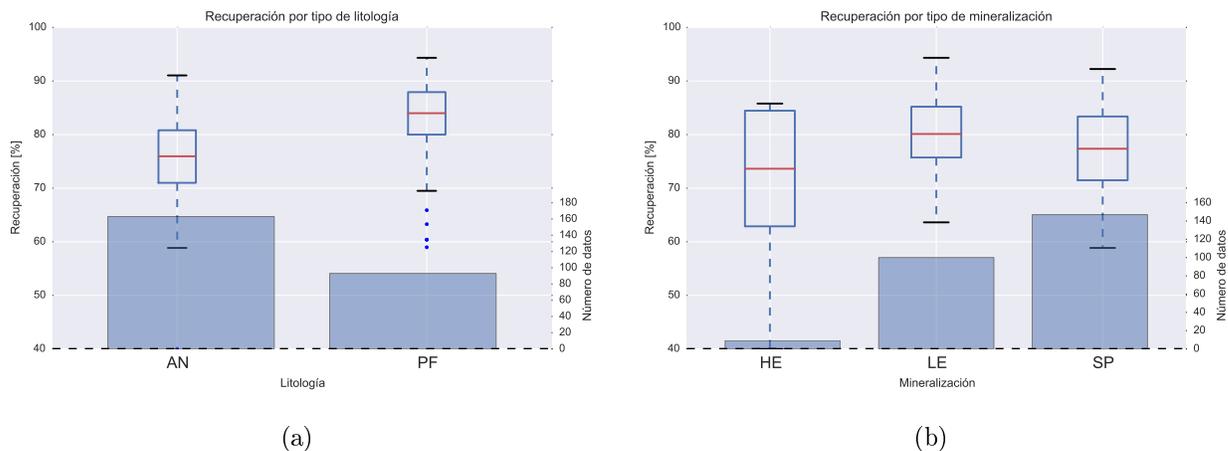


Figura C.2: Boxplots de recuperación por tipo de variable geológica en el dominio B. Metodología original. (a): Recuperación por litologías. (b): Recuperación por mineralizaciones.

C.3.3. Dominio C

Tabla C.7: Estadísticas de recuperación. Dominio C de metodología clásica.

Número de datos	574
Media	87.67 %
Desviación estándar	5.13 %
CV	0.06
Mínimo	60.73 %
25 %	85.43 %
50 %	88.70 %
75 %	91.13 %
Máximo	96.62 %
RIC	5.70 %

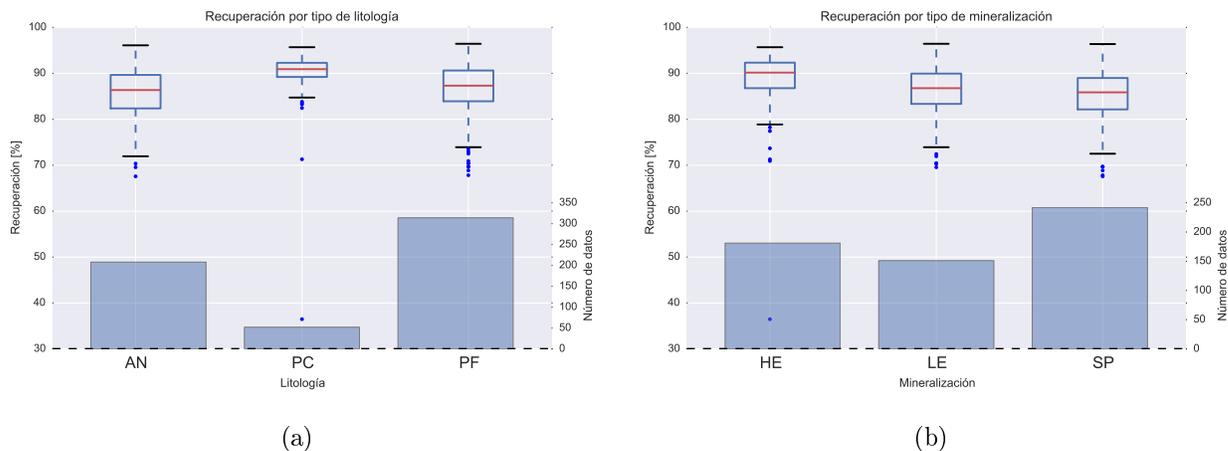


Figura C.3: Boxplots de recuperación por tipo de variable geológica en el dominio C. Metodología original. (a): Recuperación por litologías. (b): Recuperación por mineralizaciones.

C.3.4. Dominio D

Tabla C.8: Estadísticas de recuperación. Dominio D de metodología clásica.

Número de datos	573
Media	86.51 %
Desviación estándar	5.89 %
CV	0.07
Mínimo	36.47 %
25 %	83.66 %
50 %	87.17 %
75 %	90.64 %
Máximo	96.44 %
RIC	6.98 %

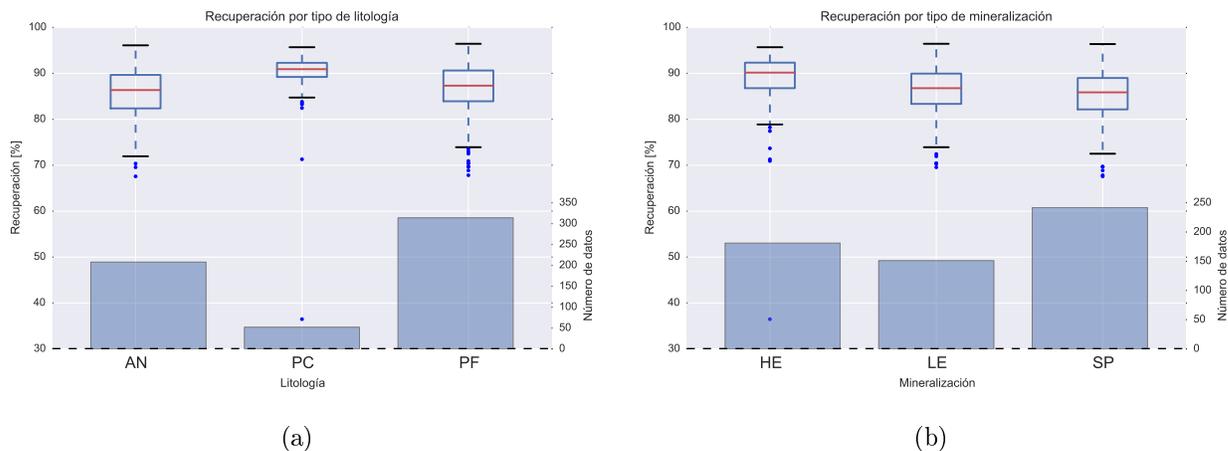


Figura C.4: Boxplots de recuperación por tipo de variable geológica en el dominio D. Metodología original. (a): Recuperación por litologías. (b): Recuperación por mineralizaciones.

C.4. Modelo predictivo

Se aplicó un modelo predictivo multilineal a través del algoritmo LARS Lasso, del cual se obtuvieron los siguientes resultados para cada dominio definido.

C.4.1. Dominio A

Tabla C.9: Resultados de regresión multilínea para dominio A, bajo diferentes parámetros.

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
(KxNa)/Al	98.787	0	0	0	0	0
3*Al/(Na+K)	0.424	0.046	0	0	0	0
Al	-1.151	0	0	0	0	-0.071
B	0	0	0	0	0	0
Bi	1.281	1.097	0.927	0.770	0.613	0.453
Cu	-1.387	-0.848	0	0	0	0
K/Al	58.997	45.580	41.469	39.431	37.392	35.235
Li	-0.272	-0.264	-0.253	-0.232	-0.211	-0.186
Na/Al	42.844	37.121	30.928	25.801	20.674	15.070
Intercepto	62.132	69.649	70.904	71.405	71.905	72.525
R2	0.395	0.387	0.382	0.374	0.360	0.342
Número de datos	149	149	149	149	149	149

C.4.2. Dominio B

Tabla C.10: Resultados de regresión multilínea para dominio B, bajo diferentes parámetros.

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
3*Al/(Na+K)	-0.359	-0.349	-0.326	-0.292	-0.258	-0.225
K/Al	-4.719	0	0	0	0	0
Na/Al	-8.735	0	0	0	0	0
Bi	0.561	0.611	0.521	0.351	0.152	0
Ag	0.108	0.038	0.021	0	0	0
Se	0.722	0.638	0.531	0.357	0.176	0
Cu	-1.220	0	0	0	0	0
Mg	4.329	0	0	0	0	0
P	-0.003	-0.002	-0.001	-0.001	0.000	0
Be	-6.891	-7.331	-6.524	-5.220	-3.914	-2.621
Li	-0.165	-0.017	0	0	0	0
Al	-4.991	-2.619	-3.043	-3.484	-3.922	-4.229
Intercepto	92.274	88.802	88.813	88.757	88.685	88.518
R2	0.578	0.566	0.561	0.546	0.520	0.486
Número de datos	256	256	256	256	256	256

C.4.3. Dominio C

Tabla C.11: Resultados de regresión multilíneal para dominio C, bajo diferentes parámetros.

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
K/Al	13.153	11.736	9.330	5.571	2.560	0
K/Mg	0.128	0	0	0	0	0
Al/Mg	-0.071	0	0	0	0	0
S	0.303	0.204	0.014	0	0	0
Cu	5.362	4.195	2.656	0	0	0
3*Al/(Na+K)	0.167	0	0	0	0	0
Mg/Al	-9.680	-8.428	-7.248	-5.282	-3.527	-1.777
Al	-3.349	-2.430	-2.597	-2.520	-2.043	-1.432
Intercepto	79.389	81.170	83.556	86.588	87.219	87.601
R2	0.454	0.446	0.418	0.336	0.253	0.140
Número de datos	574	574	574	574	574	574

C.4.4. Dominio D

Tabla C.12: Resultados de regresión multilíneal para dominio D, bajo diferentes parámetros.

	s=0	s=0.01	s=0.025	s=0.05	s=0.075	s=0.1
K/Al	0.063	0	0	0	0	0
K/Mg	-0.070	0.079	0.016	0	0	0
Al/Mg	0.151	0.006	0	0	0	0
S	0.424	0.211	0.013	0	0	0
Cu	5.619	4.530	2.873	0.367	0	0
3*Al/(Na+K)	-0.288	-0.095	-0.067	0	0	0
Mg	2.819	0	0	0	0	0
Ga	1.652	0	0	0	0	0
P	-0.002	0.000	0	0	0	0
V	-0.157	-0.063	-0.044	-0.015	0	0
Al	-3.495	-0.906	-1.331	-1.513	-0.764	0
Be	-8.804	-9.908	-10.112	-8.304	-6.444	-4.025
Intercepto	84.984	86.595	88.590	89.710	89.150	88.322
R2	0.497	0.466	0.427	0.336	0.248	0.143
Número de datos	573	573	573	573	573	573

C.5. Conclusiones

- A diferencia de la metodología propuesta en este trabajo, la definición de los dominios depende netamente de la interpretación de las variables geológicas presentes en la base de datos y no a un proceso no supervisado de PCA+Clustering.

- Sin embargo, usando esta metodología es fácil catalogar datos nuevos a los dominios definidos, ya que consiste en básicamente en identificar su asociación geológica.
- Esta metodología entrega coeficientes de determinación relativamente bajos en cuanto a los modelos predictivos, además de incorporar una gran cantidad de variables, lo que incrementa el sobre-ajuste e incorpora variables que en muchos casos pueden no estar presentes en datos nuevos.