



UNIVERSIDAD
DE ATACAMA

FACULTAD DE INGENIERÍA
DEPARTAMENTO DE GEOLOGÍA

**APLICACIÓN DE MÉTODOS DE MACHINE LEARNING PARA LA
EXPLORACIÓN GEOQUÍMICA DE TIERRAS RARAS EN YACIMIENTOS TIPO
IOCG, DISTRITOS CANDELARIA-PUNTA DEL COBRE Y MANTOVERDE,
REGIÓN DE ATACAMA, CHILE.**

Osciel Antonio Morales Madariaga

Copiapó, Chile 2023.



UNIVERSIDAD **DE ATACAMA**

FACULTAD DE INGENIERÍA
DEPARTAMENTO DE GEOLOGÍA

**APLICACIÓN DE MÉTODOS DE MACHINE LEARNING PARA LA
EXPLORACIÓN GEOQUÍMICA DE TIERRAS RARAS EN YACIMIENTOS TIPO
IOCG, DISTRITOS CANDELARIA–PUNTA DEL COBRE Y MANTOVERDE,
REGIÓN DE ATACAMA, CHILE.**

Trabajo de titulación presentado en conformidad a los requisitos para obtener el título de Geólogo.

Profesor Guía: Dr. Jaime Araya Vargas

Osciel Antonio Morales Madariaga

Copiapó, Chile 2023.

Dedicatoria

Antes de todo quiero agradecer a mis padres, por darme una infancia llena de buenos momentos y felicidad, y haberme enseñado a luchar por mis metas, entregarme las herramientas para salir adelante. A mis hermanos, quienes desde que tengo memoria siempre han estado a mi lado, entregándome su cariño, acuerdos y desacuerdos. A mis amigos, quienes me han brindado su amistad y buenos momentos vividos. Además, quiero dedicar este trabajo a aquellas personas que, aunque ya no están con nosotros o con quienes no he tenido la oportunidad de volver a ver, dejaron una huella imborrable en mi vida. Su presencia y enseñanzas, incluso en ausencia, han sido una gran motivación y fuente de inspiración para mí en todo momento.

Agradecimientos

Quiero expresar mi más profundo agradecimiento a todas las personas que han sido parte de mi formación académica. Han sido años llenos de desafíos y aprendizajes, y no podría haber llegado hasta aquí sin la ayuda de tantas personas valiosas.

En especial, quiero agradecer al Dr. Jaime Araya por ser mi profesor guía durante el desarrollo de mi tesis. Su orientación, sabiduría y apoyo fueron fundamentales en este proceso, y estoy muy agradecido por todo lo que ha hecho por mí.

También quiero agradecer a mis compañeros, quienes me brindaron información y su apoyo en todo momento. Compartir experiencias y conocimientos con ustedes fue una experiencia enriquecedora y me ayudó a crecer tanto a nivel personal como profesional.

Finalmente, quiero agradecer a todos aquellos que, de alguna u otra forma, me apoyaron en este camino. Sus palabras de aliento, sus consejos y su amistad fueron vitales en momentos difíciles. Les estaré eternamente agradecido por ser parte de mi vida y por haberme ayudado a alcanzar esta meta.

ÍNDICE GENERAL

CAPÍTULO I INTRODUCCIÓN	1
1.1. Presentación General y contexto del problema de investigación	1
1.2. Objetivos	3
1.2.1. Objetivo General.....	3
1.2.2. Objetivos Específicos	3
1.3. Hipótesis.....	4
1.4. Ubicación y accesos	4
CAPÍTULO 2: MARCO TEÓRICO	5
2.1. <i>Knowledge Discovery in Databases</i>	5
2.2. Inteligencia Artificial	7
2.3. <i>Data Mining</i>	8
2.3.1. Conceptos claves.....	8
2.3.2. Ruido.....	8
2.3.3. Sesgo	10
2.3.4. Varianza	10
2.4. <i>Machine Learning</i>	11
2.4.1. Generalidades.....	11
2.4.2. Historia.....	13
2.4.3. Enfoques de aprendizaje	15
2.4.4. <i>Underfitting y Overfitting</i>	18
2.4.5. ALGORITMOS	21
2.4.6. Métricas de <i>clustering</i>	24
2.5. Análisis de Componentes Principales (PCA)	31
2.6. <i>Machine learning</i> aplicado a la exploración mineral	32
2.6.1. <i>Machine learning</i> aplicado a la exploración de depósitos tipo IOCG	32
2.6.2 <i>Machine learning</i> aplicado a la exploración de depósitos de REE.....	34
2.7. TIERRAS RARAS	36
2.7.1. Comportamiento químico	38
2.7.2. Depósitos.....	42
2.7.3. Mineralogía	45
2.7.4. Aplicaciones.....	48

2.7.5. Mercado Mundial.....	49
2.7.6. Tierras raras en Chile	50
2.8. Yacimientos tipo IOCG.....	52
2.8.1 Génesis.....	53
2.8.2 Alteración.....	54
2.8.3 Distribución temporal	57
2.8.4 Tectónica.....	57
2.8.5 Relación genética depósitos IOA e IOCG	59
CAPÍTULO 3: MARCO GEOLÓGICO	61
3.1. Marco geológico regional.....	61
3.1.1. Geología regional de la Cordillera de la Costa.....	61
3.1.2. Geología Estructural	63
3.1.3. Cinturón de Hierro Chileno	66
3.1.4. Depósitos IOCG en la región de Atacama	67
3.1.5. Evolución y relación genética de los depósitos IOCG e IOA de la región de Atacama	70
3.1.6. REE en Atacama	71
3.2. Geología Local	73
3.2.1. Distrito Candelaria–Punta del Cobre	73
3.2.2. Distrito Mantoverde	81
CAPÍTULO 4: METODOLOGÍA	86
4.1. Descripción general de la metodología	86
4.2. Herramientas	88
4.2.1. <i>Anaconda</i>	88
4.2.2. Lenguaje de programación <i>Python</i>	88
4.2.3. Librerías	89
4.3. Base de Datos	90
4.3.1. Control de Calidad	90
4.4. Validación de Algoritmos	91
4.5. Análisis de datos con algoritmos de <i>Machine Learning</i>	91
4.5.1 Preprocesamiento	93
4.5.2 Análisis exploratorio de la base de datos	93
4.6. Análisis de Componentes Principales	95

4.7. Algoritmos utilizados y su validación	95
4.7.1. Agrupamiento Jerárquico	95
4.7.2. <i>K-means</i>	95
CAPÍTULO 5: RESULTADOS.....	101
5.1 Estadística básica de concentraciones geoquímicas	101
5.2 Análisis de Componentes Principales	105
5.3 Agrupamiento Jerárquico	108
5.4 <i>K-means</i>	113
CAPÍTULO 6: DISCUSIÓN.....	117
6.1 Elección de cantidad de clústeres para Agrupamiento Jerárquico y <i>K-means</i>	117
6.2 Validez de los resultados	118
6.3 Análisis de resultados	120
6.3.1 Análisis del Agrupamiento Jerárquico	121
6.3.2. Análisis de <i>K-means</i>	126
6.4 Correlación de resultados	131
6.4.1 Comparación Agrupamiento Jerárquico vs <i>K-means</i>	131
6.4.2 Comparación entre PCA y matriz de correlación	133
6.4.3 Asociaciones de Elementos.....	133
6.4.4 Agrupamiento utilizando dos clústeres para <i>K-means</i> y Agrupamiento Jerárquico	134
6.5 Interpretación Geológica	137
6.6 Implicancias para la exploración	142
CAPÍTULO 7: CONCLUSIÓN.....	145
REFERENCIAS.....	148
ANEXOS.....	202
ANEXO 1.- Conjunto de datos de estudio.	202
ANEXO 2.- Código	204
ANEXO 3.- Boxplots restantes de Agrupamiento Jerárquico.....	205
ANEXO 4.- Boxplots restantes de <i>K-means</i>	208
ANEXO 5.- Tabla concentraciones promedio de REE en los clústeres con respecto a los valores de la corteza.....	211
ANEXO 6.- <i>Boxplots</i> para dos clústeres	212

Índice de figuras

Figura 1.1. Imagen satelital del centro de la región del área de estudio.....	4
Figura 2.1. Una visión general de los pasos que componen el proceso KDD..	6
Figura 2.2. Tipos de inteligencia artificial	12
Figura 2.3. Esquemas de grandes áreas del <i>Machine Learning</i>	16
Figura 2.4. Ejemplo de Aprendizaje Supervisado.	19
Figura 2.5. Ejemplos gráficos de subajuste (<i>underfitting</i>) y sobreajuste (<i>overfitting</i>).....	20
Figura 2.6. Ejemplo de dendrograma con cinco clústeres	21
Figura 2.7. Tabla periódica de elementos.	37
Figura 2.8. Radios de los iones trivalentes de los elementos de tierras raras.....	40
Figura 2.9. Producción total de óxidos de REE	49
Figura 2.10. Tendencia histórica de la producción de REE en el mundo y en China.....	50
Figura 2.11. Representación esquemática de los dos modelos propuestos para el origen de los depósitos de IOCG.	55
Figura 2.12. Arquitecturas y orígenes hidrotermales alternativos para sistemas IOCG.....	56
Figura 2.13. Distribución temporal de los grandes depósitos de IOCG	58
Figura 2.14. Diagrama esquemático que muestra el entorno tectónico y litosférico de los depósitos de IOCG	59
Figura 3.1. Mapa geológico para el segmento de la Cordillera de la Costa	64
Figura 3.2. Mapa de localización de los depósitos de IOCG, IOA y pórfidos formados	69
Figura 3.3. Columna estratigráfica del distrito Candelaria-Punta del Cobre	77
Figura 3.4. Secuencia paragenética de la mineralización de IOCG en el distrito de Candelaria-Punta del Cobre.....	78
Figura 3.5. Mapa geológico esquemático del área alrededor de Copiapó y Chañaral, norte de Chile.....	82
Figura 4.1. Diagrama de trabajo.....	87
Figura 4.2. Distribución de las muestras recopiladas de los distritos Candelaria-Punta del Cobre y Mantoverde	88
Figura 4.3. Diagrama de trabajo del análisis de datos	92
Figura 4.4. Ejemplo de una matriz de correlación.	94

Figura 4.5. Ejemplo de la varianza explicado por cada componente principal	95
Figura 4.6. Ejemplo de mapa de calor PCA.....	96
Figura 4.7. Ejemplo de dendrograma jerárquico de variables	96
Figura 4.8. Ejemplo de dendrograma jerárquico de clústeres obtenidos.....	98
Figura 4.9. Ejemplos de métricas para determinar el número de clústeres.....	100
Figura 5.1. Histogramas de frecuencia de REE.	104
Figura 5.2. Matriz de correlación ordenada.	106
Figura 5.3. Mapa de calor PCA.....	107
Figura 5.4. Curva de varianza acumulada explicada.....	108
Figura 5.5. Dendrograma de Agrupación Jerárquica para las variables).	109
Figura 5.6. Dendrograma de las muestras.....	110
Figura 5.7. Distribución espacial de los clústeres de Agrupamiento Jerárquico de las muestras en el Distrito Candelaria-Punta del Cobre.....	111
Figura 5.8. Distribución espacial de los clústeres de Agrupamiento Jerárquico en el Distrito Mantoverde	112
Figura 5.9. Métricas para determinar el número óptimo de clústeres.	114
Figura 5.10 Distribución espacial de los clústeres de <i>K-means</i> de las muestras en el Distrito Candelaria-Punta del Cobre.....	115
Figura 5.11. Distribución espacial de los clústeres de <i>K-means</i> (en el Distrito Mantoverde	116
Figura 6.1. <i>Boxplots</i> de los clústeres obtenidos por Agrupamiento Jerárquico de las muestras.....	124
Figura 6.2. <i>Boxplots</i> de los clústeres obtenidos por <i>K-means</i> de las muestras.. ...	129
Figura 6.3. Distribución espacial para dos clústeres en el distrito Candelaria-Punta del Cobre	135
Figura 6.4. Distribución espacial de dos clústeres obtenidos mediante el Agrupamiento Jerárquico y <i>K-means</i> en el distrito Mantoverde	136

Índice de Tablas

Tabla 2.1. Concentración de REE en la corteza vs concentración en Proyectos de Exploración y Explotación,.....	41
Tabla 2.2. Minerales que presentan REE con un alto potencial económico.	47
Tabla 2.3. Principales usos de REE.....	48
Tabla 4.1. Métricas de evaluación de <i>clustering</i> para Agrupamiento Jerárquico.	97
Tabla 4.2. Ejemplos de métricas de rendimiento de las propiedades deseables.	100
Tabla 5.1. Estadística básica de la base de datos de elementos geoquímicos.	102
Tabla 5.2. Métricas número de clústeres para Agrupamiento Jerárquico	109
Tabla 6.1. Métricas de evaluación de <i>clustering</i> para Agrupamiento Jerárquico y <i>K-means</i> para dos y cinco clústeres.	118
Tabla 6.2. Métricas de rendimiento de las propiedades deseables para <i>K-means</i> y Agrupamiento Jerárquico vs La.....	120
Tabla 6.3. Métricas de rendimiento de las propiedades deseables para <i>K-means</i> y Agrupamiento Jerárquico vs Dy.	120
Tabla 6.4 Características de los cinco clústeres de Agrupamiento Jerárquico de las muestras.....	125
Tabla 6.5. Características de los cinco clústeres obtenida mediante el análisis de <i>K-means</i>	130
Tabla 6.6 Comparación de clústeres para cada método	132

RESUMEN

En esta tesis se utilizaron algoritmos de *Machine Learning* no supervisados, específicamente *K-means* y Agrupamiento Jerárquico, en conjunto con Análisis de Componentes Principales, para buscar patrones ocultos en la distribución de elementos de Tierras Raras (REE) en los depósitos tipo IOCG de los distritos mineros Candelaria-Punta del Cobre y Mantoverde. La base de datos analizada (81 muestras) se obtuvo de la literatura, y se procesó utilizando técnicas estadísticas y de minería de datos para identificar grupos similares de elementos. Los resultados obtenidos se analizaron para establecer su relación con la geología local. Se evaluó también la capacidad de ambos algoritmos para detectar patrones de agrupamiento en los datos, y la calidad de estos agrupamientos, empleando distintas métricas. Ambos algoritmos consistentemente agruparon a las REE en dos conjuntos: un grupo que reúne a algunas REE ligeras (La, Ce y Nd) junto con los óxidos Fe_2O_3 y MgO ; y un grupo que reúne a todas las REE pesadas (Gd, Dy, Ho, Er, Tm, Yb, y Lu) con algunas REE ligeras (Pr, Sm, Eu) y Na_2O . Los resultados obtenidos demostraron que tanto *K-means* como Agrupamiento Jerárquico son eficaces para identificar agrupamientos de elementos que son geoquímicamente coherentes, tanto por su afinidad química como por su relación con asociaciones de alteración y mineralización descritos en los distritos estudiados y otros depósitos tipo IOCG. Las asociaciones de elementos obtenidos no son concluyentes en cuanto a los minerales que hospedarían la mineralización de REE, por lo cual se interpreta la presencia de diferentes minerales portadores de REE. En base a lo obtenido, se concluye que los algoritmos utilizados son una herramienta útil para identificar asociaciones de elementos relacionados a la presencia de REE, por lo cual pueden ser de gran ayuda para identificar guías de exploración de estos elementos.

PALABRAS CLAVES: Machine Learning, IOCG, Tierras Raras, Distrito Candelaria-Punta del Cobre, Distrito Mantoverde

ABSTRACT

This thesis aimed to test the use of unsupervised Machine Learning algorithms, specifically K-means and Hierarchical Clustering in conjunction with Principal Component Analysis, in order to find hidden patterns in the distribution of Rare Earth Elements (REE) in IOCG-type deposits present in the Candelaria-Punta del Cobre and Mantoverde mining districts. The analyzed database comprises 81 samples compiled from the literature and processed using statistical and data mining techniques to identify similar groups of elements and their relationship to the local geology. The ability of algorithms to detect clustering, as well as the quality of clustering, was evaluated using different metrics. Both algorithms consistently grouped REE in two clusters: one cluster composed by some Light Rare Earth Elements (LREE) (La, Ce y Nd), Fe_2O_3 and MgO ; and a second cluster including all Heavy Rare Earth Elements (HREE) (Gd, Dy, Ho, Er, Tm, Yb, y Lu), some LREE (Pr, Sm, Eu) and Na_2O . Both clusters are coherent with expected chemical affinities and mineral associations described in IOCG deposits, indicating that these results are geologically meaningful. The obtained results are not conclusive about the minerals that host REE, suggesting that the presence of REE can be related to multiple sources. Altogether, the obtained results suggest that both K-means and Hierarchical Clustering can identify meaningful geochemical patterns, which can be used to better understand the distribution of REE in the studied mining districts.

KEY WORDS: Machine Learning, IOCG, Rare Earths Elements, Candelaria-Punta del Cobre, Mantoverde.

CAPÍTULO I

INTRODUCCIÓN

1.1. Presentación General y contexto del problema de investigación

La Comisión Chilena de Energía Nuclear (CCHEN) en conjunto a la Empresa Nacional de Minería (ENAMI) llevaron a cabo un proyecto durante los años 1995 y 2003, una investigación geológica que buscaba la presencia de elementos llamados Tierras Raras (REE, por su sigla en inglés) en la Cordillera de la Costa en las regiones III y IV. En este marco definieron tres prospectos en las cercanías de la localidad de Diego de Almagro, III región: Sierra Áspera, Cerro Carmen y Veracruz. De estos, el prospecto Cerro Carmen fue el más prometedor por su tonelaje y leyes (Comisión Chilena del Cobre [COCHILCO], 2016).

También existen depósitos de óxidos de hierro (principalmente magnetita) con altos contenido de REE. Estos depósitos pueden ser de dos subtipos óxidos de hierro-cobre-oro (IOCG) y magnetita apatito (IOA). Los minerales presentes en estos son el apatito, monacita, allanita, xenotima y fluorocarbonatos (COCHILCO, 2016). En la región de Atacama entre los depósitos del primer grupo IOCG se destacan Candelaria, Punta del Cobre y Mantoverde. Ejemplos del segundo grupo IOA son Romeral, Algarrobo, Los Colorados, Cerro Negro Norte y Cerro Imán, además de los prospectos mencionados anteriormente.

A pesar de la importancia de REE aún se sabe poco de las asociaciones minerales, y la presencia de estos elementos es difícil de identificar de manera macroscópica. Estos elementos no son menos abundantes que el plomo o el oro, pero es poco común encontrarlos en las concentraciones suficientes para que la extracción sea económicamente viable (COCHILCO, 2016). Debido al aumento del uso de REE en el mundo actual, resulta importante identificar metodologías que permitan acelerar el proceso exploratorio de estos recursos. Dentro de estas metodologías el *Machine Learning* asoma como principal candidato a liderar un cambio disruptivo en las geociencias, y la implementación de estos algoritmos pueden ayudar a incidir en nuevas guías de exploración. La integración de un enfoque de sistemas minerales con *machine learning* ha permitido el desarrollo de una serie de modelos conceptuales y se han propuesto nuevos

métodos para la cartografía de prospección mineral (p.ej., Rodríguez-Galiano et al., 2014; Costa et al., 2019; Daviran et al., 2021). El éxito de la exploración basada en este enfoque conducirá finalmente a nuevas estrategias de exploración adecuadas para descubrir nuevos yacimientos minerales (Rieuwers y Caté, 2019).

En el área de las exploraciones de yacimientos, precisamente en tipo IOCG, se han desarrollado distintos modelos exploratorios empleando algoritmos de *machine learning*. Entre los cuales destacan los métodos no supervisados *Principal Component Analysis* (PCA), *Hierarchical clustering*, *Self-Organizing Feature Map* (SOM) y *K-means* y supervisados *Random Forrest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN) y *Decision Tree* (DT) (Sarparandeh y Hezarkhani, 2016; Dmitrijeva et al., 2019b; Zaremotlagh et al., 2016; Abdolmaleki et al., 2020). Estos modelos resultaron tener éxito en la exploración de yacimientos. Mediante la integración de datos de teledetección espacial con datos geológicos usando clasificadores SOM y SVM se puede generar un mapa de prospección mineral concordante con estudios geológicos anteriores (Abdolmaleki et al., 2020). El método RF mediante la composición elemental de la magnetita fue capaz de distinguir entre los depósitos de IOCG y IOA (Hong et al., 2021). También se han realizado estudios de aplicabilidad de estos métodos en la exploración de REE. Se agruparon muestras basadas en el contenido de REE usando el algoritmo SOM (Sarparandeh y Hezarkhani, 2016), se han propuesto técnicas de exploración mediante varios métodos de clasificación donde el método de DT resultó ser el más adecuado (Zaremotlagh y Hezarkhani, 2016), y por medio del uso de KNN se propuso un enfoque de visualización de tendencias geoquímicas (Zaremotlagh y Hezarkhani, 2016).

Estudios recientes indican que yacimientos tipo hierro-cobre-oro (IOCG) en el distrito Candelaria-Punta del Cobre y Mantoverde contienen importantes créditos de REE (Marschik, 1996; Marschik et al., 2000; Marschik y Fontboté 2001a, 2001b; Mathur et al., 2002; Rieger et al, 2010a; Del Real et al., 2018). Los yacimientos de Candelaria, Carola y Trinidad contienen localmente pirrotita, esfalerita, elevadas concentraciones de LREE y trazas de arsenopirita. La allanita es el único mineral portador de REE identificado al microscopio (Marschik y Fontboté, 2001b, Mathur et al., 2002). Las concentraciones elevadas de REE y de Zn se dan principalmente cerca del contacto del Miembro Geraldo-

Negro con el Miembro Algarrobo (Marschik et al., 2000). Para explicar la génesis del yacimiento de Candelaria se han utilizado modelos en los que intervienen tanto fluidos magmáticos como no magmáticos (Ullrich et al., 1999; Marschik et al., 2000; Marschik y Fontboté, 2001a).

Estos estudios indican que la mineralización de REE se relaciona a una fuente de magma parental primitivo modificado por fluidos de subducción para el batolito de Copiapó. Los patrones de las muestras de las principales unidades intrusivas del complejo de Copiapó son similares y muestran un moderado enriquecimiento en LREE. Las rocas mozodioríticas y monzoníticas cuarzosas tienden a tener contenidos de LREE elevados (Marschik et al., 2003). Los valores normalizados de las condritas REE para las diferentes fases intrusivas del batolito, sugieren un magma diferente o una ruta de cristalización distinta que implica la segregación de plagioclasa durante el ascenso del magma (Del Real et al., 2018). La mineralización de REE en el distrito Mantoverde esta hospedada principalmente en rocas alteradas o mineralizadas. Las concentraciones elevadas de LREE ocurren localmente, son atribuibles a la redistribución de LREE por fluidos hidrotermales, pero no se presenta evidencia de que hayan sido aportados por fuentes externas. Parte de los LREE están alojados en allanita o apatita, ambos minerales están presentes solo en pequeñas cantidades (Rieger et al., 2009; Rieger et al., 2010b).

1.2. Objetivos

1.2.1. Objetivo General

Evaluar la aplicabilidad y efectividad de algoritmos de *machine learning* para predecir variaciones espaciales en las concentraciones de REE en rocas de los distritos Candelaria-Punta del Cobre y Mantoverde con mineralización tipo IOCG.

1.2.2. Objetivos Específicos

1. Construir modelos de predicción de patrones de distribución de REE.
2. Obtener patrones de distribución de REE en datos geoquímicos disponibles para las áreas de estudio.
3. Interpretar la naturaleza de los factores geológicos (petrología, mineralización, meteorización) que controlan los patrones obtenidos para la distribución de REE.
4. Evaluar el rendimiento de los algoritmos empleados.

1.3. Hipótesis

Como hipótesis se plantea que el uso de algoritmos de *Machine Learning* permitirá identificar patrones en la distribución espacial de las concentraciones de REE en ambientes con mineralización del tipo IOCG en el Distrito Candelaria-Punta del Cobre y en el distrito Mantoverde. Por lo anterior, se espera que empleando estos algoritmos será posible pronosticar zonas con mayor concentración de REE, y por lo tanto contribuir a la exploración de estos elementos.

1.4. Ubicación y accesos

Los distritos Candelaria-Punta del Cobre y Mantoverde se encuentran en el norte de Chile, región de Atacama. La Figura 1.1 muestra la ubicación de ambos distritos.

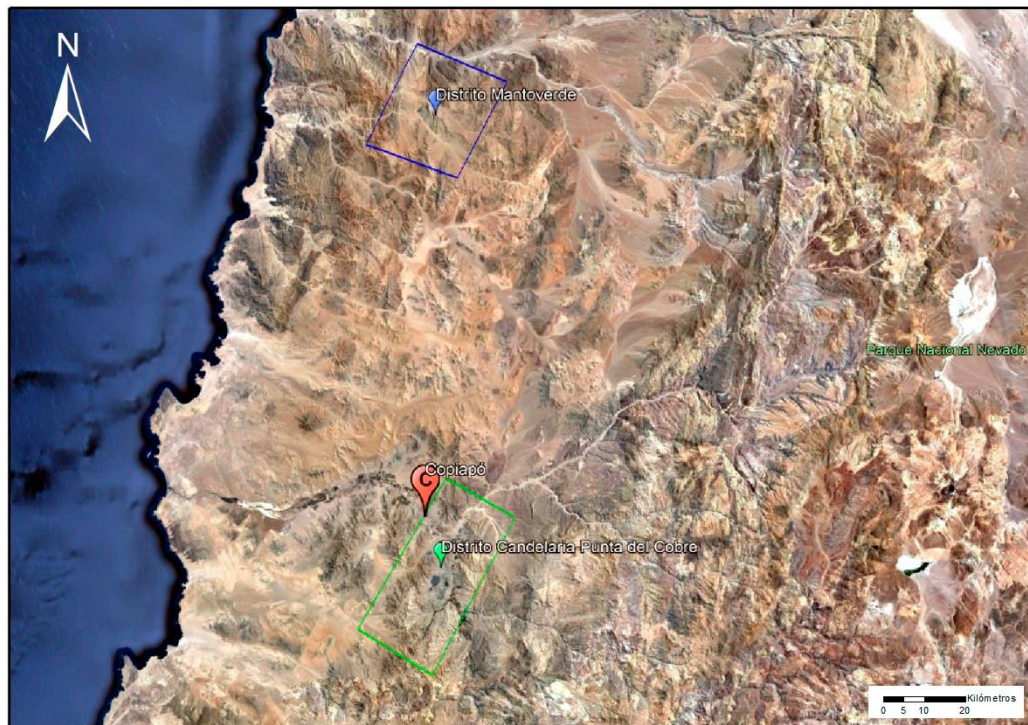


Figura 1.1. Imagen satelital del centro de la región de Atacama. El Punto C indica la ubicación de Copiapó. El rectángulo verde corresponde al distrito Candelaria Punta del Cobre. El rectángulo azul corresponde al distrito Mantoverde..

CAPÍTULO 2: MARCO TEÓRICO

En este capítulo se describe los conceptos utilizados para la elaboración de esta tesis. En la primera parte se muestra el proceso conocido como *Knowledge Discovery in Databases* (KDD) y sus etapas. Posteriormente se describe en *Machine Learning* a los algoritmos empleados.

2.1. *Knowledge Discovery in Databases*

En la literatura actual se puede encontrar un gran número de definiciones acerca de la extracción de conocimiento KDD. Se puede definir como proceso no-trivial de la inteligencia artificial de rápido crecimiento, que busca identificar patrones desconocidos, válidos, nuevos, potencialmente útiles y comprensibles dentro de los datos, mediante la combinación de *machine learning*, estadística, bases de datos y visualización de datos (Fayyad y Simoudis, 1997; Fayyad et al., 1996b, 1996c; Han et al., 2011).

El proceso de KDD se compone por 9 pasos básicos (Brachman y Anand, 1996):

1. Desarrollar una comprensión del contexto de la aplicación y del conocimiento previo relevante, e identificar el objetivo del KDD.
2. Crear un conjunto de datos objetivo: seleccionar un conjunto de datos acordes con el paso anterior.
3. Limpieza y preprocesamiento de datos: eliminación de ruido, recopilación de información necesaria para el modelado, decidir cómo tratar datos faltantes, contabilización de la información de la secuencia temporal y los cambios conocidos.
4. Reducción y proyección de datos: se buscan características útiles para representar los datos objetivos. Se utilizan métodos de reducción de la dimensionalidad o de transformación para reducir el número efectivo de variables consideradas o para encontrar representaciones invariantes de los datos.
5. Encontrar los métodos concretos de extracción de datos afines a los objetivos del proceso de KDD (paso 1) por ejemplo: resumen, clasificación, regresión, agrupación, etc.

6. Elección de los algoritmos de minería de datos: Seleccionar los métodos que se utilizarán para buscar patrones en los datos. Esto incluye decidir qué modelos y parámetros pueden ser apropiados y hacer coincidir un método de minería de datos particular con los criterios generales del proceso KDD.
7. *Data mining*: búsqueda de patrones de comportamiento de interés que sea representativa en la base de datos seleccionada.
8. Interpretación de los patrones extraídos, de ser necesario volver a los pasos 1-7 para iterar. Visualización de los patrones encontrados, o la visualización de los datos del modelo.
9. Consolidar el conocimiento encontrado: añadir el conocimiento a otro sistema para filtrarlo, o documentarlos. Comprobación y resolución de posibles conflictos con los conocimientos creados (o extraídos) previamente.

El proceso global de KDD (figura 2.1) incluye la evaluación y posible interpretación de los patrones "extraídos" para determinar qué patrones pueden considerarse nuevos "conocimientos" (Fayyad et al., 1996a).

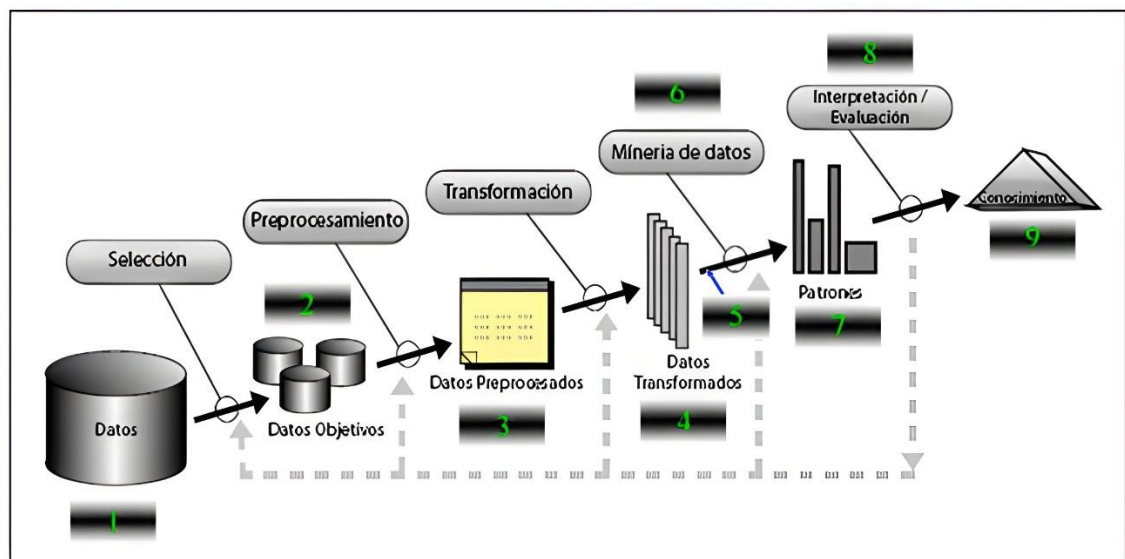


Figura 2.1. Una visión general de los pasos que componen el proceso KDD. Los números indican a qué paso pertenecen (modificado de Fayyad et al., 1996a).

2.2. Inteligencia Artificial

La inteligencia artificial (IA) es en informática una disciplina enfocada en la creación de artilugios artificiales (e.g. máquinas como un ordenador), para realizar tareas propias de la inteligencia humana (Poole et al., 1998; Benko y Lánvi, 2009; Ramírez, 2020).

Marvin Minsky, cofundador del Laboratorio de Inteligencia Artificial y Ciencias de la Computación (CSAIL) del Instituto de Tecnología de Massachusetts (MIT por sus siglas en inglés); lo definió como "una disciplina que se encarga de crear máquinas programadas que sean capaces de hacer cosas que requieren la misma inteligencia que si fuesen hechas por humanos" (McCarthy et al., 2006).

A lo largo de la historia se han seguido cuatro enfoques, los cuales suelen estar clasificados en dos grupos (Russell y Norvig, 2004):

- Enfoques centrados en los humanos: sistemas que piensan como humanos (Bellman, 1978; Haugeland, 1985) y sistemas que actúan como humanos (Kurzweil, 1990; Rich y Knight, 1991). El enfoque centrado en el comportamiento humano debe ser una ciencia empírica, que incluye hipótesis y confirmaciones mediante experimentos.
- Enfoques centrados en la racionalidad: sistemas que piensan racionalmente (Charniak y McDermott, 1985; Winston, 1992) y sistemas que actúan racionalmente (Poole et al., 1998; Nilsson, 1965, 1998). El enfoque racional implica una combinación de matemáticas e ingeniería.

La inteligencia no es una dimensión única, sino un espacio profusamente estructurado de capacidades diversas para procesar la información. Algunas (como razonar) se suelen describir como “inteligentes”. Otras (como la visión), no. Pero todas entrañan competencias psicológicas (como la percepción, la asociación, la predicción, la planificación, el control motor) que permiten a los seres humanos y demás animales alcanzar sus objetivos (Boden, 2017).

2.3. Data Mining

El *Data Mining* es el conjunto de técnicas y tecnologías que permiten extraer información y encontrar patrones repetitivos de interés con conocimientos útiles y comprensibles desde grandes bases de datos de manera automática o semiautomática que expliquen el comportamiento de estos datos (Han et al., 2011, 2012; Bramer, 2007). Antes de adentrarse al *machine learning*, se deben definir algunos conceptos claves de la *Data Mining*.

2.3.1. Conceptos claves

2.3.1.1. Conjunto de datos (Dataset)

El conjunto de datos son una parte integral de los sistemas de predicción. Son los datos históricos usados para entrenar un sistema o modelo de detección de patrones. Este compuesto por instancias, e instancias de factores, características o propiedades (Torralba y Efros, 2011; Gónzales, s.f). El conjunto de datos engloba a los conjuntos de entrenamiento, validación y prueba. Es importante que el conjunto de datos tenga un volumen suficiente como para generar resultados estadísticamente significativos, y a la vez, que sea representativo del conjunto de datos global (Santos, 2022).

2.3.1.2. Datos de entrenamiento

Los datos de entrenamiento son aquellos datos etiquetados usados para entrenar un modelo o algoritmo, con el fin de que aprenda de dichos conjuntos, ajuste los parámetros y aumente la precisión. Suele ser el conjunto más grande (Batista et al., 2004; Zhu et al., 2012). La calidad del modelo de *machine learning* va a ser directamente proporcional a la calidad de los datos. Es por ello por lo que las labores de preprocesamiento, limpieza y depuración consumen un gran porcentaje del tiempo de trabajo (Santos, 2022; Cracknell, 2014).

2.3.1.3. Datos de prueba

Los datos de prueba son los datos reservados para evaluar la generalización del modelo. Evalúa que tan bien funciona el modelo generado a partir de los datos de entrenamiento. En otras palabras, si las repuestas son acertadas o no para un caso nuevo (Howden, 1975; Mitchell, 1999; Bhavsar y Ganatra, 2012).

2.3.1.4. Datos de validación

Los datos de validación ayudan a evaluar o validar los hiperparámetros del modelo, previniendo el sobreajuste o subajuste (Breck et al., 2019).

2.3.1.5. Función objetivo

La función objetivo es aquella función que será optimizada dadas las limitaciones o restricciones determinadas y con variables que necesitan ser minimizadas o maximizadas mediante programación lineal o no lineal (Saaty y Gass, 1954; Gass y Saaty, 1955; Bezdek et al., 1981).

2.3.1.6. Hiperparámetros

Los Hiperparámetros son los valores de las configuraciones utilizadas durante el proceso de entrenamiento. Son utilizados para parametrizar el proceso de instanciación del modelo (Rozada, 2019; Wang y Gong, 2018). Estos valores generalmente no se obtienen de los datos, sino que son indicados por el científico de datos. No se usan para modelar el modelo directamente, influyen en la capacidad y características de aprendizaje del modelo (Mackay, 1996). Ejemplos de hiperparámetros pueden ser el número de vecinos de *K-Nearest Neighbors*, la profundidad máxima de un árbol de decisión, etc. El valor óptimo de un hiperparámetro no puede ser conocido a priori para un problema dado. Para ello se utilizan valores genéricos, reglas genéricas, los valores que han funcionado anteriormente en problemas similares. La validación cruzada es una buena opción para buscar los hiperparámetro (Rodríguez, 2021).

2.3.1.7. Etiquetado de datos

El etiquetado de datos se denomina al proceso de identificación de las propiedades de los datos (documentos, texto, imágenes, números, etc.) y asignándole un nombre a dichas propiedades. Los datos pueden estar etiquetados y no etiquetados (Evans y Eysers, 2012; Cracknell, 2014). Se dice que un conjunto de datos está etiquetado cuando éstos tengan asociados una etiqueta, y que serán datos sin etiquetar cuando no lo tengan. Cabe destacar que un conjunto de datos puede ser “no etiquetado” por qué la etiqueta no existe, pero no

haya sido aún añadida a los datos o porque no existe (InteractiveChaos, s.f; Bruening y Waterman, 2010).

2.3.2. Ruido

El ruido es cualquier anomalía, error aleatorio de varianza de una variable no medida, no deseada en los datos y, debido al ruido, la clase puede ser más difícil de aprender (Han et al, 2011; Alpaydin, 2020). Según Alpaydin (2020) hay varias interpretaciones del ruido:

- Puede haber imprecisión en el registro de los atributos de entrada, lo que puede desplazar los puntos de datos en el espacio de entrada.
- Puede haber errores en el etiquetado de los puntos de datos, que pueden reetiquetar los casos positivos como negativos y viceversa. Esto se denomina a veces *ruido del profesor*.
- Puede haber atributos adicionales, que no hemos tenido en cuenta, que afecten a la etiqueta de una instancia. Estos atributos pueden estar *ocultos* o *latentes*, es decir, pueden ser inobservables. El efecto de estos atributos descuidados se modela, pues, como un componente aleatorio y se incluye en el "ruido".

2.3.3. Sesgo

En *machine learning* el término sesgo (*bias* en inglés) fue introducido por Mitchell (1980), el cual refiere al error persistente o sistemático que se espera que cometa el algoritmo de aprendizaje cuando se entrena con conjuntos de entrenamientos de tamaño m (Dietterich y Kong, 1995). En otras palabras, se puede definir como la diferencia entre la predicción esperada del modelo con respecto a los valores verdaderos observados. El sesgo habitualmente presenta una relación inversa con la complejidad de los modelos, a mayor complejidad del modelo utilizado es de esperar un menor sesgo (Rodríguez, 2019). Un sesgo elevado puede hacer que un algoritmo pase por alto las relaciones relevantes entre las características y los resultados objetivo (subajuste o *underfitting*) (Campesato, 2020). El sesgo en los sistemas de IA proviene de diversas fuentes, desde el algoritmo y los atributos de entrada elegidos hasta las correlaciones ocultas en los datos de entrenamiento (Roselli et al, 2019).

2.3.4. Varianza

Es la cantidad en que la estimación de la función objetivo cambiará si se utiliza diferentes datos de entrenamiento (Gonzales, 2018b). La varianza indica el grado de dispersión de los datos (Han et al, 2003). El error debido a la varianza es la variabilidad de la predicción de un modelo para un punto de datos determinado (Campesato, 2020).

La varianza puede ser baja apuntando pequeños cambios en la estimación de la función objetivo con cambios en el conjunto de datos de entrenamiento, Por el contrario, una alta varianza sugiere grandes cambios en la estimación de la función objetivo con cambios en el conjunto de datos de entrenamiento (Gonzales, 2018b).

2.4. *Machine Learning*

2.4.1. Generalidades

El *Machine Learning* (ML) es un subconjunto de la IA que permite a un sistema aprender de una muestra de datos mediante la programación de algoritmos informáticos que son capaces de emular la inteligencia humana (El Naqa y Murphy, 2015; Alpaydin, 2020). El ML ha sido aplicado en campos tan diversos como lo son la biología (Skolnick et al., 2021), la ingeniería aeroespacial (Ao et al., 2010), reconocimiento de patrones (Bishop, 2006), la visión por ordenador (Apolloni et al., 2005), servicios financieros (Györfi et al., 2012), deportes (Musa et al., 2019), astronomía (Ball y Brunner, 2010), aplicaciones biomédicas y médicas (Waljee y Higgins, 2010). La figura 2.2 muestra la relación entre la IA y sus diferentes subcampos.

El ML se basa en ideas de diferentes disciplinas como la probabilidad, la estadística, la informática, la teoría de la información y reconocimientos de patrones, la psicología, la teoría del control, la filosofía y las aplicaciones de minería de datos (Michie et al. 1994; Ripley 1996; Mitchell, 1997; Bishop, 2006; Hastie et al. 2009; Alpaydin, 2020). La principal característica del ML es que su aprendizaje es basado en datos empíricos y puede modelar un fenómeno sin conocer las reglas físicas que lo gobiernan y buscan automatizar ciertos procesos en diferentes áreas (Jorquera, 2019; Rojas, 2020). Los algoritmos van aprendiendo a medida que son ejecutados (Jofré, 2020).

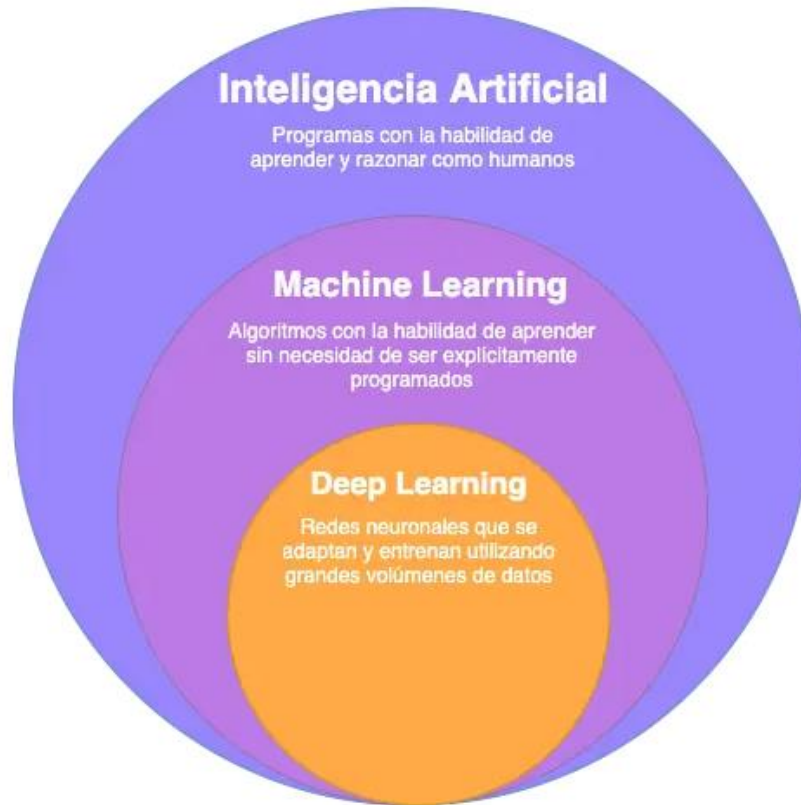


Figura 2.2. Tipos de inteligencia artificial (extraído de Ventajas del *Machine Learning* y Deep Learning, 2019).

La esencia del ML es dotar a los ordenadores la capacidad de adaptar sus estructuras de decisión, basándose en las características de los datos observados, y generar predicciones válidas y objetivas (Schölkopf 2003). Para la investigación científica, el reconocimiento de patrones es el proceso de descubrir, mediante métodos estadísticos automatizados o semiautomatizados, patrones útiles dentro de los datos (Kotsiantis et al., 2007).

Existen dos enfoques lógicos para la inferencia de datos: el razonamiento deductivo y el inductivo (Feng y Michie 1994; Sivia 1996; Gubbins 2004). Los enfoques deductivos de la inferencia suelen implicar a un investigador (o analista) que desarrolla una hipótesis (o modelo) sobre algún fenómeno natural. Posteriormente, se utilizan los datos de observación para aceptar o rechazar la hipótesis. Por su parte, los enfoques inductivos de la inferencia utilizan inicialmente los datos/observaciones disponibles para identificar patrones. A continuación, los patrones aprendidos se emplean para desarrollar una serie de hipótesis plausibles (Sivia 1996; Burl et al. 1998).

2.4.1.1. Entrenamiento de modelo de *Machine Learning*

El entrenamiento de un modelo de ML consiste en proporcionar datos de entrenamiento de los cuales aprende un algoritmo de ML. Es la fase posterior a la selección y tratamiento de datos, busca encontrar los hiperparámetros que mejor predigan resultados a partir de los datos y tenga un mejor comportamiento (Géron, 2019; AWS Amazon, s.f).

2.4.1.2. Diferencia entre modelo y algoritmo

Antes de continuar, se debe hacer una diferenciación entre el concepto de modelo y algoritmo. Existen varias acepciones del concepto de modelo. El modelo puede considerarse como una representación de la realidad, explicación de un fenómeno, paradigma; patrón para seguir o muestra para conocer algo; guía de acción; arquetipo; prototipo; uno entre una serie de objetos similares; explicar un proceso a o abstracción, un conjunto de elementos esenciales o los supuestos teóricos de un sistema social (Caracheo, 2002; Aguilera, 2000).

Para este trabajo un modelo es un conjunto de parámetros y de instrucciones que nos dice como usar los parámetros (Hacking, 1983). A su vez, un algoritmo es un conjunto ordenado de operaciones sistemáticas que permite calcular los valores óptimos para los parámetros de nuestro modelo (Gareth et al., 2013). Por ende, un modelo puede tener uno o más algoritmos.

Los algoritmos de aprendizaje automático (ALM) son herramientas computacionales que proporcionan a los analistas métodos estadísticamente sólidos para la inferencia inductiva de datos (MacKay 2003; Bousquet et al. 2004). En otras palabras, se busca entrenar un modelo que tenga la capacidad de predecir los parámetros objetivos definidos (Rojas, 2020).

Con esta definición el modelo de ML puede ser entrenado con un conjunto de datos y, proporcionarle un ALM que pueda averiguar y obtener información de esos datos (IBM, s.f.).

2.4.2. Historia

El origen del campo de la Inteligencia Artificial se puede atribuir a Alan Turing que en 1950 creó el “Test de Turing” para determinar si una máquina era realmente inteligente.

Para completar el test, una máquina tenía que ser capaz de engañar a un humano haciéndole creer que era humana en lugar de una máquina (French, 2000). En la historia moderna, Arthur Samuel acuñó el término "*machine learning*" definiéndolo como "un campo de estudio que da a los ordenadores la capacidad de aprender sin ser programados explícitamente" y demostró que los ordenadores podían ser programados para aprender a jugar a las damas (Samuel, 1959). Esto fue seguido por el desarrollo del perceptrón como una de las primeras arquitecturas de redes neuronales en 1958 basándose en ideas sobre el funcionamiento del sistema nervioso humano (Rosenblatt, 1958). Mitchell (1997), utilizando léxico de ciencias de la computación presento el *machine learning* como "se dice que un programa de ordenador aprende de la experiencia (E) con respecto a alguna clase de tareas (T) y medida de rendimiento (P), si su rendimiento en tareas en T , medido por P , mejora con la experiencia E ".

A principios de la década de los 60's se diseñaron y probaron sistemas de reconocimiento (Widrow, 1961; Braverman, 1961). También se desarrollaron enfoques deterministas (Vapnik y Lerner, 1963; Vapnik y Chervonenkis, 1964; Yakubovich, 1963, 1965, 1966) y se sugirió el planteamiento probabilístico del problema del ML (Aizerman et al, 1964a, b). La creación del algoritmo "Nearest Neighbor" marca el inicio del reconocimiento de patrones en computadoras (Cover y Hart, 1967). Se propuso un enfoque estadístico en base a las formulaciones estadísticas de las comunicaciones y la teoría de la decisión (Marill y Green, 1960; Widrow, 1959).

El comienzo de la primera década del siglo XXI se convirtió en un punto de inflexión en la historia del lenguaje de signos, lo que se explica por tres tendencias sincrónicas que, en conjunto, produjeron un efecto sinérgico apreciable: la primera tendencia es la de los *Big Data*, reducir el coste de la informática y la memoria en paralelo, y el desarrollo de nuevos algoritmos de aprendizaje automático profundo (*deep learning*); cuya cantidad de datos se hizo tan grande que los nuevos enfoques se hicieron realidad por necesidad práctica y no por curiosidad de los científicos (Fradkov, 2020).

2.4.3. Enfoques de aprendizaje

Dentro de las técnicas de ML podemos encontrar distintos enfoques bien diferenciados como lo son el aprendizaje supervisado, el aprendizaje no supervisado, el aprendizaje reforzado, y el más desconocido y menos empleado el aprendizaje semi-supervisado. La figura 2.3 presenta un esquema de los principales enfoques de aprendizaje del ML, así como sus aplicaciones.

2.4.3.1. Aprendizaje supervisado

El aprendizaje supervisado es aquel que busca la relación existente entre una variable de entrada (*input*) y una variable de salida (*output*) deseada. A los valores de *input* se le asignan etiquetas con el fin de clasificarlas mediante clases y minimizar el error o pérdida, el cual corresponde a la diferencia entre el valor predicho y el valor real deseado (Cracknell, 2014; Marsland, 2009). A continuación, se espera que, con la información entregada, el algoritmo sea capaz de relacionar las variables con las clases presentes en los datos de entrenamiento y predecir el *output* (Kuncheva 2004; Hastie et al. 2009).

Por ejemplo, un algoritmo es capaz de clasificar imágenes entre gatos, perros y otros. Para ello se deben etiquetar mediante clases. Para una imagen de gato se asignan un valor de 1, para una imagen de perro un valor de 2, y para otros un valor de 3. Entonces a partir de un conjunto de datos de entrenamiento se ingresa la imagen (*input*) de un gato y se define su salida como 1, así sucesivamente para cada dato de entrenamiento (figura 2.4a). Después de muchos ejemplos el algoritmo será capaz de clasificar de manera correcta, incluso valores que no haya visto antes (figura 2.4b). En otras palabras, mediante observación, el modelo fue capaz de generalizar el conocimiento.

Entre los algoritmos de aprendizaje supervisado se encuentran la regresión simbólica, árboles de decisión, *support vector machine* (SVMs), entre otros (Nasteski, 2017). Sus aplicaciones más comunes incluyen la clasificación y las regresiones (Goleman et al, citado por Jofré, 2020).

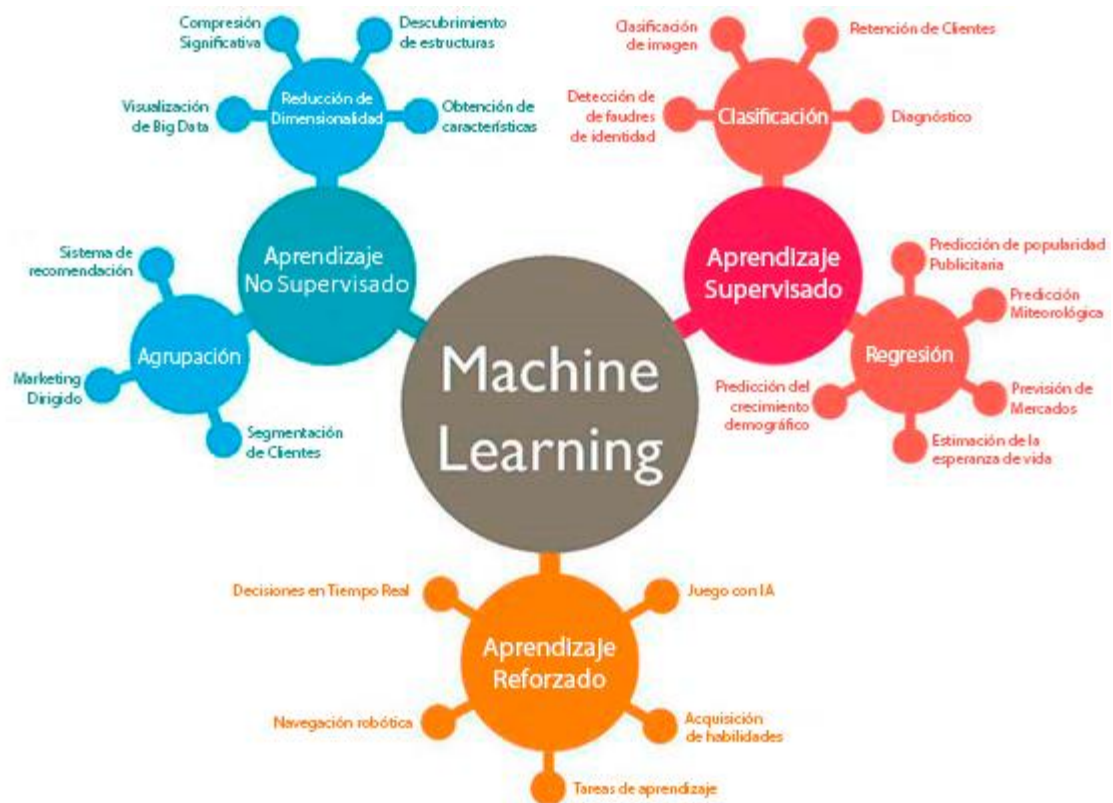


Figura 2.3. Esquemas de grandes áreas del *Machine Learning* (Yalalov, 2022).

2.4.3.2. Aprendizaje no supervisado

Este enfoque genera conocimiento únicamente con los datos de entradas (*input*), a diferencia del aprendizaje supervisado que requiere también una variable de salida (*output*) definido. El objetivo del aprendizaje no supervisado no es minimizar el error, sino encontrar características similares entre los datos, para realizar una clasificación correcta usando patrones coherentes que por otros métodos convencionales no se pueden lograr por su complejidad (Barlow, 1989; Kuncheva 2004; Hastie et al. 2009). Por ejemplo, es frecuente el uso de este enfoque por parte de las grandes empresas de comercio electrónico en las campañas de marketing las cuales segmentan a los consumidores con características o comportamientos similares como lo pueden ser el grupo etario, género, gustos, entre otros, con la finalidad de llevar sus productos a nichos en específico.

Al no estar entrenado, se suele mencionar que los algoritmos de aprendizaje no supervisado están “basado en datos” (Ripley, 1996). Son utilizados principalmente cuando se tiene un gran volumen de datos sin clasificación (Khanum et al., 2015). Los algoritmos

de Aprendizaje No Supervisado son útiles para identificar y visualizar clústeres (grupos naturales) dentro de los datos (Rohwer et al. 1994; Ripley 1996; Witten y Frank 2005). Las muestras dentro de un clúster determinado deberían ser similares entre sí, mientras que las muestras asociadas a clústeres diferentes no deberían serlo (Kuncheva 2004; Xu y Wunsch 2005; Marsland 2009). Ejemplos de algoritmos de aprendizaje no supervisado son el *clustering*, reducción de dimensión, la visualización, *Hierarchical Clustering* (Agrupación Jerárquica), *K-means*, Self-Organising Maps (Ripley 1996; Witten y Frank 2005). En este trabajo se utilizó el término “*clustering*” para referirse a la técnica de agrupamiento de datos en conjuntos homogéneos. Todos los algoritmos utilizados en esta tesis son del tipo aprendizaje no supervisado.

2.4.3.3. Reforzado

Son modelos que tienen una subestructura muy compleja, y se entrenan a sí mismos por repetición, a lo largo de millones de iteraciones, intentando realizar de la mejor forma una tarea en particular. Ejemplos de este enfoque son los modelos entrenados por la empresa DeepMind para jugar juegos de mesa como el ajedrez o el “Go”, en los cuales los modelos eran capaces de anticipar con exactitud mediante predicciones las jugadas de jugadores profesionales (Holcomb et al., 2018; Mitchell y Lumsden, 2018). En el aprendizaje reforzado, el programa no posee un conjunto de datos objetivos a los que llegar, sino que a través de ensayo y error mejora su comportamiento frente a un problema específico (Gray, 2019). Estos algoritmos están inspirados en la psicología conductivista, cuya ocupación es determinar qué acciones debe escoger un agente de software en un entorno dado con el fin de maximizar alguna noción de “recompensa” o premio acumulado (Williams, 1987; Bertsekas y Tsitsiklis, 1996).

2.4.3.4. Semi-supervisado

El aprendizaje semi-supervisado combina el aprendizaje supervisado y no supervisado (Chapelle et al., 2006; Zhu 2008). Utiliza datos de entrenamiento tanto etiquetados como no etiquetados para realizar determinadas tareas de aprendizaje, generalmente una pequeña cantidad de datos etiquetados junto a una gran cantidad de datos no etiquetados (Van Engelen y Hoos, 2020). Este enfoque permite construir mejores clasificadores, requiere menos esfuerzo humano y proporciona una mayor precisión (Zhu, 2008). Los recientes avances en la investigación del ML han demostrado que los métodos de

aprendizaje semi-supervisado pueden resolver problemas de clasificación similares con datos mucho menos acotados que los métodos de aprendizaje supervisado (Huo et al., 2019).

2.4.4. *Underfitting y Overfitting*

En este apartado se va a tratar el tema del *underfitting* y *overfitting*, dos conceptos importantes en los modelos de ML. Es muy común que los científicos de datos, en especial aquellos que no tienen mucha experiencia, entrenen modelos que, a la hora de predecir datos nuevos, el modelo tiene un mal comportamiento. El *underfitting* y *overfitting* son peligros constantes y omnipresentes en el ML (Bashir et al., 2020). El problema que surge al momento de crear un modelo es que se busca que no “memorice” los datos de entrenamiento, si no que aprenda la relación real. En este escenario existe una técnica de la ciencia de datos bien establecida que permite desarrollar o escoger el modelo óptimo: la validación (Bashir et al., 2020).

2.4.4.1. *Underfitting*

El *underfitting* (subajuste) se refiere a un modelo que ha memorizado demasiado el conjunto de datos de entrenamiento y no puede generalizar o encajar bien en un conjunto de datos nuevos, esto ocurre cuando el modelo creado es muy simple (Guyon y Yao, 1999; Andrews, 2018; Gonzales, 2018a). El *underfitting* ocurre cuando un algoritmo aprende los detalles y el ruido en el conjunto de entrenamiento en la medida en que impacta negativamente el rendimiento del modelo en un nuevo conjunto de datos, por lo que el clasificador resultante tendrá una alta tasa de error tanto en el conjunto de entrenamiento como en nuevos datos, siendo incapaz de encontrar la relación entre las variables de entrada y salida con precisión (Lawrence et al., 1997; IBM, 2021).

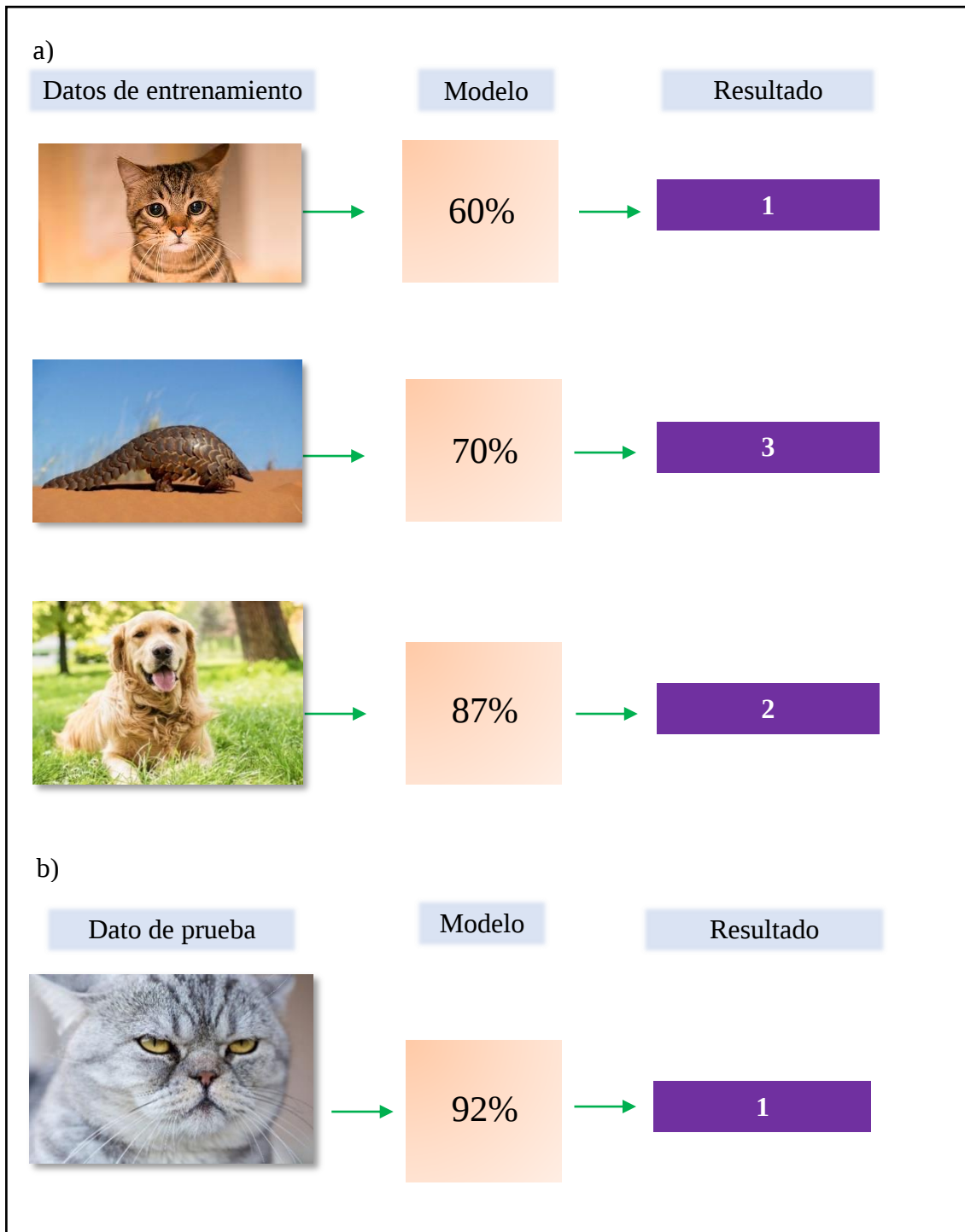


Figura 2.4. En a) se usa el conjunto de entrenamiento para aplicar el modelo. El porcentaje hace referencia al grado de acierto que logra el modelo durante el entrenamiento a medida que se ingresan los datos. En b) el modelo entrenado alcanza un acierto del 92%, siendo capaz de predecir el resultado de un dato desconocido. (Elaboración propia).

2.4.4.2. Overfitting

El *overfitting* (sobreajuste) se produce cuando un algoritmo reduce el error mediante la memorización de ejemplos de entrenamiento, con características ruidosas o irrelevantes, en lugar de aprender la verdadera relación general entre X e Y (Piotrowski y Napiorkowski, 2013; Krueger et al., 2017; Zhang et al., 2017).

En el caso del *overfitting*, el algoritmo está demasiado adaptado a los datos de entrenamiento que le son enseñados y será incapaz de generalizar nuevos datos de entrada que le son desconocidos. Cuando el modelo está sobreentrenado, el algoritmo solo considerará válidos los datos idénticos a los del conjunto de entrenamiento incluyendo el ruido (Chan et al., 2011; Panchal et al., 2011; Domingos, 2000). Esto significa que el algoritmo, en un momento determinado del período de entrenamiento, ya no mejora su capacidad para resolver el problema, sino que simplemente empieza a aprender alguna regularidad aleatoria contenida en el conjunto de patrones de entrenamiento (Jabbar y Khan, 2015). En la figura 2.5 se puede observar un ejemplo gráfico sobre el *underfitting* y el *overfitting*.

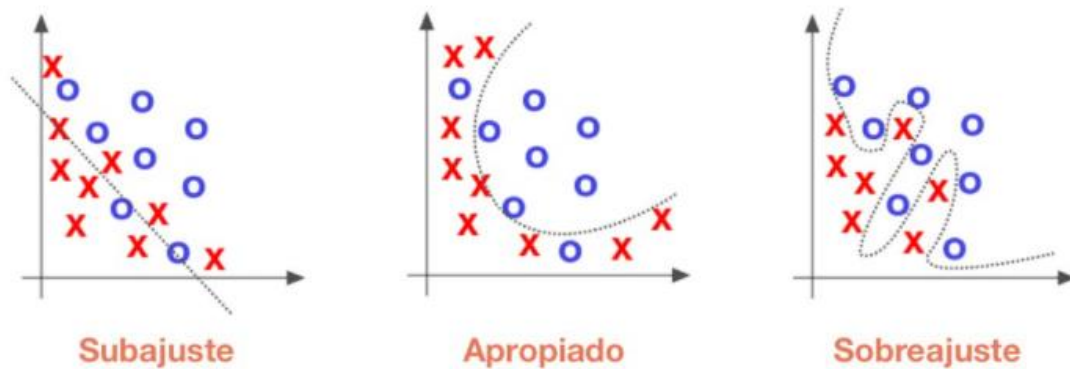


Figura 2.5. Ejemplos gráficos de subajuste (*underfitting*) y sobreajuste (*overfitting*). En el gráfico izquierdo se aprecia que la línea (modelo) no es capaz de explicar la distribución de los datos, por lo que el modelo está mal ajustado (*underfitting*). Por su parte, el gráfico de la derecha muestra una línea que ajusta bien todos los puntos del gráfico, pero también modela el ruido de los datos y valores atípicos, por lo cual el modelo sobreajusta los datos (*overfitting*). El gráfico del medio muestra un ejemplo de un ajuste apropiado, donde la línea describe razonablemente bien la distribución de los puntos, y sería eventualmente capaz de aprender y generalizar manteniendo el equilibrio entre el bias y la varianza. Extraído de *AprendeIA* [Imagen], Gonzales (2018a), <https://aprendeia.com/sobreajuste-y-subajuste-en-machine-learning/>.

2.4.5. ALGORITMOS

A continuación, se procede a detallar algunos de los algoritmos que se emplearon.

2.4.5.1. Agrupamiento Jerárquico

El agrupamiento jerárquico es una técnica de agrupamiento que busca construir una jerarquía de clústeres. Es decir, cada punto de datos comienza siendo su propio clúster y luego se van uniendo en clústeres más grandes según su similitud, hasta que se llega a un solo clúster que contiene todos los datos. Los algoritmos de *clustering* jerárquico generan incrementalmente un dendrograma o árbol 1D que representa las similitudes entre clústeres (Jain y Dubes, 1988). Los dendrogramas pueden construirse utilizando métodos aglomerativos donde cada punto de datos comienza como un clúster individual y se van uniendo en clústeres más grandes según su similitud (abajo a arriba), o divisivos donde todos los datos comienzan en un solo clúster y se van separando en clústeres más pequeños (de arriba a abajo) (Cracknell, 2014; Stahl y Sallis, 2012). Las disimilitudes entre las muestras dentro de los clústeres se especifican mediante los niveles de ramificación de un dendrograma (Figura 2.6).

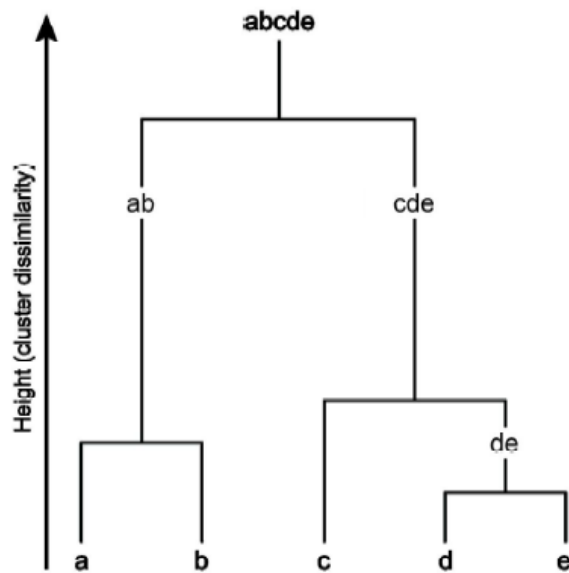


Figura 2.6. Ejemplo de dendrograma con cinco clústeres como resultado de la agrupación jerárquica; los clústeres similares residen en la misma rama del dendrograma. Extraído de Cracknell (2014).

El agrupamiento jerárquico utiliza medidas de similitud como la distancia euclidiana, la distancia de Manhattan o la correlación de Pearson para fusionar dos clústeres con las menores disimilitudes medidas (Ripley, 1996). Los métodos de enlace simple, completo y medio definen la disimilitud entre los clústeres fusionados y tienen diferentes criterios para unir clústeres en un nivel determinado. Por ejemplo, el método de enlace simple une dos clústeres si se pueden encontrar enlaces que unan pares de muestras con disimilitudes inferiores al nivel especificado. El método de enlace completo fusiona dos clústeres si todas las muestras de un clúster están a una distancia mínima de las muestras del otro clúster, mientras que el método de enlace medio considera que la disimilitud combinada entre dos clústeres es la media de todas las disimilitudes existentes entre las muestras de cada clúster (Jain y Dubes, 1988; Ripley 1996; Murty y Devi, 2011).

2.4.5.2. K-means

El algoritmo de *clustering K-means* es un método numérico, basada en la medida euclidiana, no determinista de clasificación no supervisada, y ampliamente utilizado para agrupar objetos en k grupos (Cracknell, 2014; Na et al., 2010; Xu y Wunsch II, 2005; Alpaydin, 2020). Clasifica los objetos en k *clústeres* diferentes a través de iteración, convergiendo a un mínimo local, logrando que los *clústeres* generados sean compactos e independientes (Cheung, 2003). Es simple y muy rápido, por lo que, en muchas aplicaciones prácticas, el método ha demostrado ser una forma muy eficaz que puede producir buenos resultados de *clustering* (Na et al., 2010). La cuantificación de vectores es una de las aplicaciones del *clustering*, pero éste también se utiliza para el preprocesamiento antes de una etapa posterior de clasificación o regresión (Alpaydin, 2020).

El algoritmo *K-means* funciona de la siguiente manera. Primero, *K-means* genera aleatoriamente un centroide que distribuye n objetos de un conjunto de datos D en k clústeres, $(C_1, \dots, C_k)$. A continuación, ajusta de forma iterativa la ubicación de este centro en el espacio de la variable euclidiana con respecto a las muestras que se consideran asociadas a este clúster. La calidad del clúster C_i puede calcularse mediante la suma del error cuadrático entre todos los objetos de C_i y el centroide x_i según la ecuación (1.1) (Ripley 1996; Witten Frank 2005; Marsland 2009; Han et al. 2012). El algoritmo consta

de dos fases distintas (Fahim et al., 2006): La primera fase selecciona k centros de forma aleatoria, donde el valor k se fija de antemano. La siguiente fase consiste en llevar cada objeto de datos al centro más cercano.

Suponiendo que el objeto objetivo es x , x_i indica la media del clúster C_i , la función objetivo se define como sigue (Na et al., 2010):

$$E = \sum_{i=1}^k \sum_{x \in C_i} |x - x_i|^2 \quad (1.1)$$

E es la Suma del Error al Cuadrado (SSE por sus siglas en ingles) de todos los objetos de la base de datos. La distancia de la función de criterio es la distancia euclidiana, que se utiliza para determinar la distancia más cercana entre cada objeto de datos y el centro del clúster. La distancia euclidiana entre un vector $x = (x_1, x_2, \dots, x_n)$ y otro vector $y = (y_1, y_2, \dots, y_n)$, la distancia euclidiana $d(x_i, y_i)$ se puede obtener como sigue:

$$d(x_i, y_i) = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2} \quad (1.2)$$

El proceso del algoritmo *K-means* es el siguiente (Mackay, 2003; Na et al., 2010):

Entrada (input):

Número de clústeres deseados, k , y una base de datos $D = \{d_1, d_2, \dots, d_n\}$ que contiene n objetos de datos.

Output:

Un conjunto de k clústeres

Pasos:

- 1) Elegir el número de clústeres k .
- 2) Seleccionar aleatoriamente los centroides.
- 3) Asignar los Clústeres y evaluar la calidad de los clústeres, con la distancia euclidiana y función objetivo.
- 4) Recalcular los centroides.
- 5) Repetir los pasos 3 y 4 hasta que el algoritmo converja hasta un error seleccionado.

El rendimiento de un algoritmo de *clustering* puede verse afectado por el valor elegido de K . Por lo tanto, en lugar de utilizar un único K predefinido, se podría adoptar un conjunto de valores (Pham et al., 2005).

2.4.6. Métricas de *clustering*

Es esencial determinar el número óptimo de clústeres en los métodos de agrupación de los algoritmos de aprendizaje no supervisado para comprender la estructura de los datos y mejorar la interpretación de los resultados. Se utilizan varias métricas para determinar el número adecuado de clústeres, como el método del "Codo", Coeficiente de *Silhouette*, *Gap Statistics*, Distorsión, Índice Davies Bouldin, Calinski-Harabasz y Cophenética (Ketchen y Shook, 1996; Milligan y Cooper, 1985; Sun et al., 2008). En este trabajo, se utilizaron varias de estas métricas para el algoritmo *K-means* y el Agrupamiento Jerárquico. Se detallan cada una de estas métricas a continuación.

2.4.6.1. Cophenética

La Cophenética es una medida de coherencia que se utiliza para evaluar la calidad de las agrupaciones jerárquicas. Este índice mide la correlación entre la distancia original de los puntos y la distancia en el dendrograma obtenido por agrupamiento jerárquico. En términos generales, cuanto mayor sea el valor de la Cophenética, más coherente será la agrupación jerárquica (Sneath y Sokal, 1973; Kaufman y Rousseeuw, 1990).

La Cophenética se calcula mediante la comparación de las distancias originales y las distancias cophenéticas. La distancia original entre dos objetos se mide como la distancia euclidiana entre ellos. La distancia Cophenética entre dos objetos se define como la altura del nodo en el que se fusionan en la jerarquía. En este sentido, la Cophenética mide la similitud entre los árboles jerárquicos generados a partir de las distancias originales y cophenéticas (Sokal y Rohlf, 1962; Jain y Dubes, 1988).

Supongamos que se ha utilizado un método de clústeres para modelar los datos originales $\{X_i\}$ y producir un dendrograma $\{T_i\}$. Este dendrograma es un modelo simplificado en el que los datos que son "ceranos" se agrupan en un árbol jerárquico.

Para calcular la Cophenética se utiliza la fórmula:

$$c = \frac{\sum_{i < j} (x(i, j) - \bar{x})(t(i, j) - \bar{t})}{\sqrt{[\sum_{i < j} (x(i, j) - \bar{x})^2] [\sum_{i < j} (t(i, j) - \bar{t})^2]}} \quad (1.3)$$

Donde $x(i, j)$ es la distancia euclidiana entre la i -ésima y la j -ésima observación, y $t(i, j)$ la distancia dendrogramática entre los puntos modelo t_i y t_j . Esta distancia es la altura del nodo en el que estos dos puntos se unen por primera vez (Sneath y Sokal, 1973).

2.4.6.2. Davies Bouldin

La métrica de Davies Bouldin es una métrica de evaluación de calidad para el agrupamiento de datos. Esta métrica considera tanto la separación como la cohesión de los clústeres y se calcula como la media del índice de similitud de Davies-Bouldin para cada clúster en relación con el clúster más similar (Mirkin, 2012). La métrica se calcula como el promedio de la máxima similitud entre cualquier clúster y todos los demás clústeres. El valor de la métrica es mayor cuanto mayor sea la similitud entre los clústeres, lo que indica una agrupación peor. Por lo tanto, un valor bajo de Davies Bouldin Score indica una buena agrupación y un valor alto indica una agrupación pobre (Davies y Bouldin, 1979). En general, se considera que un valor de Davies Bouldin Score inferior a 1 es aceptable para una buena agrupación.

$$DB = \frac{1}{k} \sum_{i=1, i \neq j}^k \max \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \quad (1.4)$$

2.4.6.3. Calinski-Harabasz

La métrica Calinski-Harabasz es una de las medidas más utilizadas para determinar el número óptimo de clústeres en un conjunto de datos. El índice Calinski-Harabasz mide la separación entre clústeres y la cohesión dentro de los clústeres, y se calcula como la relación entre la varianza interclúster y la varianza intraclúster. La métrica Calinski-Harabasz es una medida interna, lo que significa que no requiere conocimiento previo sobre los datos y no depende de la distribución de los datos (Calinski y Harabasz, 1974; Zhang y Huang, 2008). Sin embargo, puede ser sensible a la presencia de valores atípicos y a la forma de los clústeres. Por lo tanto, es recomendable emplearlo junto con otras

medidas internas y externas para determinar el número óptimo de clústeres en un conjunto de datos (Milligan y Cooper, 1985).

Para calcular el índice Calinski-Harabasz se debe seguir una serie de pasos (Sidyakov, 2022). El primer paso consiste en calcular la dispersión intercluster o la suma de cuadrados entre grupos (BGSS, por sus siglas en ingles). La dispersión entre clústeres en Calinski-Harabasz mide la suma ponderada de las distancias al cuadrado entre los centroides de un clúster y el centroide de todo el conjunto de datos. La suma de cuadrados entre grupos se calcula como:

$$BGSS = \sum_{k=1}^K n_k \times \|C_k - C\|^2 \quad (1.5)$$

Donde n_k es el número de observaciones en el clúster k , C_k es el centroide del clúster k , C es el centroide de la base de datos, y K el número de clústeres.

El segundo paso consiste en calcular la dispersión intracluster o la suma de cuadrados dentro del grupo (WGSS, por sus siglas en ingles). La dispersión intracluster en Calinski-Harabasz mide la suma de las distancias al cuadrado entre cada observación y el centroide de este clúster. Para cada clúster k , se calcula la $WGSS_k$ como:

$$WGSS_k = \sum_{i=1}^{n_k} \|X_{ik} - C_k\|^2 \quad (1.6)$$

Donde n_k es el número de observaciones en el clúster k , X_{ik} es la i -ésima observación del clúster k , C_k es el centroide del clúster k . Luego se suma todas las sumas de cuadrados individuales dentro del grupo:

$$WGSS = \sum_{k=1}^K WGSS_k \quad (1.7)$$

Donde $WGSS_k$ es la suma de cuadrados dentro del grupo del clúster k .

Finalmente se calcula el índice Calinski-Harabasz, la cual se calcula como:

$$CH = \frac{\frac{BGSS}{K-1}}{\frac{WGSS}{N-K}} = \frac{BGSS}{WGSS} \times \frac{N-K}{K-1} \quad (1.8)$$

Donde $BGSS$ es la suma de cuadrados entre grupos (dispersión entre grupos), $WGSS$ es la suma de cuadrados dentro de los grupos (dispersión dentro de los grupos), N es el número total de observaciones y K el número total de clústeres.

2.4.6.4. *Silhouette*

El Coeficiente de *Silhouette* es una métrica utilizada para evaluar la calidad de los resultados del análisis de agrupamiento obtenidos por algoritmos de *clustering*. Este coeficiente se calcula para cada objeto y mide la similitud de los puntos dentro de un clúster en comparación con los puntos de otros clústeres (Rousseeuw, 1987). El coeficiente varía entre -1 y 1, donde un valor cercano a 1 indica una buena separación entre clústeres y un valor cercano a -1 indica una mala separación. La fórmula para calcular el coeficiente de *Silhouette* de un objeto se basa en la diferencia entre la distancia media del objeto a todos los demás objetos en su clúster y la distancia media del objeto a todos los objetos en el clúster más cercano, dividida por el máximo de estas dos distancias (Rousseeuw, 1987; Kaufman y Rousseeuw, 2009). La fórmula para calcular *Silhouette* de un objeto es:

$$s(x) = \frac{b(x) - a(x)}{\max \{a(x), b(x)\}} \quad (1.9)$$

Donde el valor de $s(x)$ puede variar de -1 y 1, $a(x)$ es la cohesión y $b(x)$ es la separación. El coeficiente de *Silhouette* para todo el agrupamiento es:

$$SC = \frac{1}{N} \sum_{i=1}^N s(x) \quad (1.10)$$

Donde N es el número de objetos en el conjunto de datos.

2.4.6.5. Método del Codo

El método del "Codo" es una técnica gráfica para determinar el número óptimo de clústeres en un análisis de agrupamiento (Thorndike, 1953). Se utiliza la suma de errores al cuadrado (SSE, por sus siglas en inglés) como indicador de rendimiento. El objetivo es encontrar el punto en la gráfica donde el descenso de la suma de las distancias cuadradas se estabiliza, lo que se asemeja a un "Codo", indicando el número óptimo de clústeres (Ketchen y Shook, 1996; Syakur et al., 2018). Los valores más pequeños indican que cada

clúster es más convergente (Liu y Deng, 2021). A menudo se utiliza junto con otras técnicas para una mayor precisión (Yuan y Yang, 2019). Se establece la función de coste utilizando la SSE.

$$J = \sum_{i=1}^k \sum_{x \in C_i} |x - C_i|^2 \quad (1.11)$$

2.4.6.6. *Gap statistic*

Gap Statistic es un algoritmo propuesto para determinar el número de clústeres de conjuntos de datos con números de clasificación desconocidos (Tibshirani et al., 2001). El método *Gap Statistic* compara el logaritmo natural de la dispersión intraclúster para diferentes valores de k (el número de clústeres) con el logaritmo natural de la dispersión esperada en un conjunto de datos aleatorio del mismo tamaño y dimensión. El número óptimo de clústeres se define como el valor k para el cual la diferencia entre la dispersión intraclúster observada y la esperada es máxima (Tibshirani et al., 2001; Xiao y Yu, 2007; Yuan y Yang, 2019). Se calcula del siguiente modo.

$$Gap_n(k) = E_n^*\{\log(W_k)\} - \log(W_k) \quad (1.12)$$

Donde E_n^* denota la expectación bajo un tamaño de muestra n desde la distribución de referencia. W_k es la suma de cuadrados de la dispersión intraclúster para k clústeres. El valor óptimo de k se determina seleccionando el valor k para el cual $Gap_n(k)$ supera el valor de $Gap_n(k) + 1$ por una cantidad mayor al error estándar.

2.4.7. Métricas de validez

Una notable desventaja de los algoritmos de aprendizaje no supervisado es su dificultad de evaluar su rendimiento, debido a que la mayoría de los modelos, si no todos, no entregan una respuesta “correcta” garantizada, por el contrario, presentan patrones ocultos de interés entre las variables del modelo (Feizollah et al., 2014; Siless et al., 2013). Para resolver este problema se emplean las métricas propuestas por Feizollah et al. (2014).

2.4.7.1. Inercia

La Inercia corresponde a la suma de las distancias al cuadrado entre cada punto y el centroide del clúster al que pertenece (Kaufman y Rousseeuw, 1990). La Inercia se puede

interpretar como una medida de la varianza total explicada por el modelo de *clustering*. En general, se busca minimizar la Inercia para obtener una agrupación óptima (Li y Rowley, 2002).

$$I = \sum_{j=1}^k \sum_{x \in C_j} \|x - \mu_j\|^2 \quad (1.13)$$

Donde k es el número de grupos, C_j es el j -ésimo grupo, μ_j es el centroide del j -ésimo grupo y $\|x - \mu_j\|^2$ es la distancia euclidiana al cuadrado entre el punto x y el centroide μ_j del grupo.

2.4.7.2. Homogeneity

Homogeneity se trata cuando en una agrupación cada clúster contiene aquellos puntos de datos que solo sean miembros de una única clase. Un valor de homogeneidad cercano a 1 indica una buena homogeneidad dentro de los clústeres, mientras que un valor cercano a 0 indica una mala homogeneidad. (Rosenber y Hirschberg, 2007; Siless et al., 2013).

$$h = 1 - \frac{H(C|K)}{H(C)} \quad (1.14)$$

Donde C representa los valores verdaderos de la clase de los datos, K representa las asignaciones de clúster hechas por el algoritmo de *clustering*, $H(C)$ es la entropía de C , y $H(C|K)$ es la entropía condicional de C dado K . En el contexto de la métrica *Homogeneity* para la evaluación de algoritmos de *clustering*, la entropía se refiere a la medida de la distribución de las clases en cada clúster.

2.4.7.3. Completeness

Completeness es una medida de evaluación para algoritmos de *clustering* que evalúa en qué medida todos los miembros de una clase se agrupan en el mismo clúster. Es decir, *Completeness* mide si todos los puntos de la misma clase están asignados al mismo clúster. Un valor de completitud cercano a 1 indica una buena asignación de clústeres, mientras que un valor cercano a 0 indica una mala asignación (Kaufman y Rousseeuw, 1990).

$$c = 1 - \frac{H(K|C)}{H(K)} \quad (1.15)$$

Donde K representa las asignaciones de clúster hechas por el algoritmo $H(K)$, y C es la división inicial.

2.4.7.4. *V-Measure*

V-Measure es una medida basada en la entropía que mide el grado de cumplimiento de los criterios de *Homogeneity* y *Completeness*. La *V-Measure* se calcula como la media armónica de las distintas puntuaciones de homogeneidad y exhaustividad (Van Rijsbergen, 1979). *V-Measure* puede utilizarse para establecer comparaciones entre conjuntos de datos (Rosenber y Hirschberg, 2007; Acock y Stavig, 1979). Donde h es *Homogeneity* y c es *Completeness*.

$$v = 2 \frac{hc}{h + c} \quad (1.16)$$

2.4.7.5. Índice de Rad Ajustado (ARI)

Índice de Rad Ajustado (ARI, por sus siglas en inglés): mide la similitud entre los clústeres formados y una asignación de clústeres previamente conocida. Calcula el porcentaje de pares de muestras que son clasificados de la misma manera por los dos agrupamientos, teniendo en cuenta todas las posibles clasificaciones (Hubert, L y Arabie, 1985). El ARI devuelve un valor entre -1 y 1, donde 1 indica una coincidencia perfecta entre los dos agrupamientos, 0 indica una agrupación aleatoria y -1 indica una falta de similitud entre los dos agrupamientos (Rand, 1971; Vinh et al., 2009). Para calcular ARI, primero se debe calcular Índice de Ajuste de Rand (RI), que mide la similitud entre dos agrupaciones.

$$RI = \frac{2(a + b)}{n(n - 1)} \quad (1.17)$$

Donde n es el número de observaciones de una muestra, a es el número de pares de observaciones con las mismas etiquetas y situadas en el mismo clúster, y b el número de observaciones con etiquetas diferentes y situadas en clústeres diferentes. Posteriormente se calcula ARI:

$$ARI = \frac{RI - R[RI]}{\max(RI) - E[RI]} \quad (1.18)$$

2.4.7.6. Información Mutua Ajustada (AMI)

Información Mutua Ajustada (AMI, por sus siglas en inglés): de igual manera que ARI, AMI es utilizada para para comparar agrupaciones. El AMI es una métrica que mide la dependencia mutua entre dos variables. A diferencia del índice de Rand ajustado, el AMI toma en cuenta el hecho de que algunas parejas de clústeres pueden aparecer por casualidad y no por una similitud real. El valor del AMI varía entre 0 y 1, donde 0 indica que las dos particiones son completamente diferentes y 1 indica una partición idéntica (Vinh et al., 2009). Para calcular AMI, primero se debe calcular la Información Mutua (MI).

$$MI(X, Y) = \frac{\sum \sum P(X, Y) \times \log(P(X, Y))}{P(X)P(Y)} \quad (1.19)$$

Donde $P(X, Y)$ es la probabilidad conjunta de que X e Y estén en el mismo grupo y $P(X)$ y $P(Y)$ son las probabilidades marginales de X e Y , respectivamente.

$$AMI = \frac{MI - E[MI]}{\max(H(U), H(V)) - E[MI]} \quad (1.20)$$

Donde MI es la información mutua entre las etiquetas verdaderas y predichas, $H(U)$ es la entropía de las etiquetas verdaderas y $H(V)$ es la entropía de las etiquetas predichas. $E[MI]$ es la información mutua esperada bajo la hipótesis nula de independencia entre las etiquetas verdaderas y predichas.

2.5. Análisis de Componentes Principales (PCA)

El análisis de componentes principales (PCA, por sus siglas en inglés) es una técnica de reducción de dimensionalidad ampliamente utilizada en el análisis de datos multivariados (Jolliffe, 2002). La idea central es transformar el conjunto original de variables en otro conjunto de nuevas variables no correlacionadas entre sí, llamado conjunto de componentes principales, la cual es la proyección lineal que conserva la variabilidad de los datos originales en un espacio de menor dimensión (Miesch, 1980; Gower y Legendre, 1986). Esta proyección lineal se define como una combinación lineal de las variables

originales que maximiza la varianza explicada en el espacio de menor dimensión. El PCA permite identificar patrones y estructuras subyacentes en un conjunto de datos al reducir la cantidad de variables originales a través de la identificación de las combinaciones lineales más importantes (Jolliffe, 2002). El PCA ha sido empleado en diversos estudios multivariantes de inferencia geológica/geocientífica (por ejemplo, Granath 1988; Gelfort 2006; Grebby et al. 2011; Yu et al. 2012; Gazley et al., 2015; Dmitrijeva et al., 2019b; Nathwani et al., 2022).

Para el cálculo de las componentes principales se considera una serie de variables ($x_1, x_2, \dots, x_p$) sobre un grupo de objetos. A partir de ellas, se busca calcular un nuevo conjunto de variables $y_1, y_2, \dots, y_p$ incorreladas entre sí, cuyas varianzas vayan decreciendo progresivamente. Cada y_j (donde $j=1, \dots, p$) es una combinación lineal de las $x_1, x_2, \dots, x_p$ originales, es decir:

$$y_j = a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jp}x_p = a'_j x \quad (1.21)$$

Siendo $a'_j = (a_{j1}, a_{j2}, \dots, a_{jp})$ un vector de constantes, y $x = \begin{bmatrix} x_1 \\ \dots \\ x_p \end{bmatrix}$

El primer componente principal se calcula eligiendo a_1 de modo que y_1 tenga la mayor varianza posible. El segundo componente principal se calcula obteniendo a_2 de modo que la variable obtenida, y_2 no esté correlada con y_1 . Esta fórmula fue tomada de Zapata (2020).

2.6. Machine learning aplicado a la exploración mineral

2.6.1. Machine learning aplicado a la exploración de depósitos tipo IOCG

En el área de exploración de depósitos IOCG, se han desarrollado distintos modelos exploratorios empleando algoritmos de ML. Zandiyyeh et al. (2016) demostró que el método *Support Vector Machine* (SVM) es eficaz en la integración de múltiples capas de información en la cartografía de prospección mineral como la alteración hidrotermal, mapeo litológico y datos geoquímicos, siendo capaz con el algoritmo de identificar yacimientos conocidos como también algunos objetivos nuevos. Dmitrijeva et al. (2019b) utilizaron análisis estadísticos multivariantes incluyendo algoritmos de *machine learning*

no supervisados (*Principal Component Analysis*, Agrupamiento Jerárquico y *K-means*) y supervisados (*Random Forest*), en un conjunto de datos de rocas para comprender e identificar las firmas geoquímicas de la mineralización. Los resultados obtenidos proporcionan una visión de la evolución de los sistemas de IOCG dominantes en magnetita a hematita, pudiendo proporcionar una guía para la exploración de depósitos de IOCG poco profundos. Utilizando un enfoque similar Adolmaleki et al (2020) estableció un procedimiento óptimo para la integración de datos de teledetección en espacial con datos geofísicos, geoquímicos y observaciones de campo de mapeo de alteración mineral (óxido de hierro e hidróxidos), para la exploración de mineralización de IOCG. Para la clasificación y el análisis de los datos utilizaron mapas autoorganizados (SOM) y SVM, con el cual se generó un mapa de prospección mineral el cual resultó concordante con los estudios geológicos anteriores. A diferencia de las investigaciones anteriores Hong et al. (2021) compararon el uso del microanalizador de sonda electrónica (EPMA) y la espectrometría de masas con plasma acoplado por láser (LA-ICP-MS), RF para clasificar los yacimientos de IOCG e IOA mediante la composición química de la magnetita. Obtuvieron como resultado que la composición elemental de la magnetita puede distinguir los depósitos de IOCG y IOA. Señalan que el rendimiento de RF basado en los datos de LA-ICP-MS es mejor que los datos de EMPA.

El ML ha demostrado ser una herramienta poderosa en la elaboración de mapas geológicos en la combinación de la Inversión del Vector de Magnetización (MVI por sus siglas en inglés) en conjunto con el algoritmo *Random Forest* en el reconocimiento de patrones de datos magnéticos obtenidos a partir de métodos aerotransportados magnetométricos, radiométricos y de topografía satelital, en zonas de alta vegetación que dificultan el número de afloramientos disponibles como es el caso de la selva amazónica (Costa et al., 2019).

Skirrow et al. (2019) desarrollaron un método de cartografía prospectiva con los componentes del sistema mineral integrando diversos datos de entrada litológicos, alteración hidrotermal, geoquímicos, y geofísicos aplicables a escalas regionales y continentales para sistemas hidrotermales y ortomagmáticos. El método no solo funciona en *brownfield* si no también en *greenfield* sin mineralización previamente conocida identificando no solo el tipo de yacimiento elegido (IOCG), sino también las variantes e

híbridos de depósitos cuya formación también está controlado por los mismos procesos geológicos o similares en el área de Tennant Creek – Mt Isa. Los resultados de los análisis del potencial de IOCG basados en el conocimiento se han validado con respecto a los depósitos de IOCG conocidos.

Debido al auge del empleo de los algoritmos de ML en geociencia, la empresa OZ Minerals en un esfuerzo por acelerar el descubrimiento de nuevos yacimientos económicos en la meseta de Mt Woods, Australia, en donde en 2001 se descubrió un yacimiento de IOCG; crearon el concurso “*Explorer Challenge*” a principios de 2019, liberando más de 2 TB de sus datos geocientíficos privados a participantes externos. Consideran que el enfoque de sistemas minerales con el aprendizaje automático ha permitido el desarrollo de una serie de modelos conceptuales para orientar una serie de tipos de depósitos y materias primas (Riuwers y Caté, 2019). El concurso de tres meses de duración concluyó el 31 de mayo de 2019, con más mil participantes de 62 países, resultando ganador Team Guru los cuales usaron un enfoque en la geoquímica, geofísica y topografía (Glesson, 2019). Aunque la empresa se reserva la información sobre los resultados de los distintos modelos.

2.6.2 *Machine learning* aplicado a la exploración de depósitos de REE

Recientemente se han realizado investigaciones de la aplicabilidad de algoritmos de ML en yacimientos minerales que contienen REE en Irán. Sarparandeh y Hezarkhani (2016) aplicaron el algoritmo de SOM, como un método no supervisado, para agrupar muestras basadas en el contenido de REEs en el yacimiento de Fe-REE de Choghart, Irán central. SOM encontró cuatro clústeres como número óptimo, por lo que el área de estudio fue subdividida en cuatro zonas, las cuales se describen como tipo fosfato, tipo albitófiro, mineral de hierro metasomático y fósforo, y tipo mineral de hierro. Siendo el tipo fosfato el más propenso a los elementos de tierras raras. Encontraron que el SOM funcionó bien y la preservación de la topología es una de sus ventajas para la exploración geoquímica. Sarparandeh y Hezarkhani (2017) señalan que los enfoques multivariables son un procedimiento más preciso. Aplicaron cinco clasificadores (supervisados) *multilayer perceptron* (MLP), el *K-Nearest Neighbor* (KNN), *Parzen windows* y SVM en el yacimiento de magnetita-apatita tipo Kiruna de Se-Chahun, Irán central. Usaron los métodos en muestras litológicas basados en REEs, P y Fe. Los clasificadores MLP, KNN,

Parzen y SVM demostraron tener un rendimiento aceptable con una tasa de clasificación correcta superior al 75%, pero MLP y KNN tuvieron mejor rendimiento en comparación a los demás clasificadores, siendo MLP el más generalizable que KNN y ser un enfoque fácil, aunque tienen en cuenta la posibilidad de sobreajuste en el método MLP. La clasificación de patrones demostró ser un enfoque apropiado para la evaluación de la distribución de REE. También Zaremotlagh y Hezarkhani (2016) propusieron una técnica de exploración utilizando varios métodos de clasificación para identificar las diferentes concentraciones de REE y sus patrones ocultos, encontrando que el método de DT fue el más adecuado. Un enfoque geoquímico basado en gráficos fue propuesto por (Zaremotlagh et al., 2016) para determinar la relación entre los objetos, empleando medidas de similitud relacionados con la composición de la roca, el patrón de REE y la ubicación geográfica del yacimiento Chogart, Irán central. Emplearon la técnica de KNN para visualizar zonas adyacentes y mostrar las tendencias geoquímicas con la finalidad de confirmar los patrones de REE en la detección de los diferentes grupos de rocas y reconocer los grupos correspondientes que sean compositiva y genéticamente similares.

Bournas et al (2019) aplicaron ML a datos geofísicos magnéticos y espectrométricos de rayos gamma aéreos enfocada en toda Burkina Faso con el objetivo de identificar nuevas áreas prospectivas de recursos minerales. Esta área se caracteriza por ser una provincia rica en oro cuyo potencial mineral está vinculado a los cinturones volcano-sedimentarios del Birimian. Además, existen una serie de yacimientos no auríferos que están representados por yacimientos polimetálicos, yacimientos de pórfido de cobre-oro y cobre-molibdeno, manganeso asociado a los sedimentos volcánicos de Tarkwa, yacimientos de kimberlita diamantífera en la zona de Barsalogho, yacimientos de hierro-titanio-vanadio asociados a gabros noríticos, mineralización laterítica de níquel-cromita-cobalto, bauxita en los sedimentos neoproterozoicos y diversas vetas y lentes mineralizadas de metales básicos. El uso de ML aplicados a los datos geofísicos mencionados ayudó a identificar nuevas áreas prospectivas para una amplia gama de recursos minerales. El área con mayor potencial es la zona más oriental y la parte sur de esta zona, que está poco explorada, parece presentar una alta prospectividad tanto para la mineralización de uranio como de REE. Los productos o mapas cartográficos obtenidos indican que Burkina Faso tiene un potencial mineral diversificado e importante no solo

para el oro, sino también para otros minerales no auríferos como uranio, metales básicos y metales estratégicos como REE.

Babaie y Davarpanah (2020), mediante el criterio de autoorganización están trabajando en un modelo aun en desarrollo de conocimiento semántico de ontología *REE_MinSys* que representa los procesos dinámicos primarios y secundarios que ocurren en una amplia gama de escalas espaciales y temporales que producen los depósitos aflorantes de REE ya sean ortomagmáticos o hidrotermales; y su geometría, tonelaje y ley. El modelo de conocimiento semántico de ontología es un lenguaje de marcado que busca codificar un documento que, junto con el texto, incorpora etiquetas o marcas que contienen información adicional acerca de la estructura del texto o su representación, según las propiedades y relaciones entre entidades que existen para un dominio en particular. Busca compartir conocimiento sobre la estructura de las cosas y en el significado de las palabras sin ninguna clase de ambigüedades. En este sistema los usuarios pueden consultar la base de datos de conocimientos para extraer hechos explícitos e inferidos en cada tipo de yacimiento de REE, por ejemplo, preguntando: "¿Qué elementos de tierras raras hay en los depósitos de fosfato de REE?", entre otros.

2.7. TIERRAS RARAS

Los elementos de tierras raras (REE por sus siglas en inglés) son un grupo de 17 elementos con propiedades químicas y físicas similares que suelen estar asociados entre sí en la corteza terrestre, pertenecientes al Grupo IIIA de la tabla periódica (Figura 2.7) (López 2017; Avendaño, 2017; Voncken, 2016; Young, 2019; Papangelakis y Moldoveanu, 2014). Además, tienden a aparecer en cualquier ocurrencia natural como un grupo en lugar de individualmente o como una combinación de unos pocos de su número (Henderson, 2013).

Entre las REE se encuentran 15 lantánidos: lantano (La), cerio (Ce), praseodimio (Pr), neodimio (Nd), prometio (Pm), samario (Sm), europio (Eu), gadolinio (Gd), terbio (Tb), disprosio (Dy), holmio (Ho), erbio (Er), tulio (Tm), iterbio (Yb) y lutecio (Lu). Dentro de las REE se consideran además escandio (Sc) e itrio (Y); estos últimos se agregan debido a sus similitudes fisicoquímicas con los lantánidos (Balcerzak, 2000; Gupta y Krishnamurthy, 2005; Emsbo et al., 2015; Jha, 2014; Long et al., 2012; Simandl, 2014).

El itrio se asocia con los lantánidos debido a que su radio iónico es muy similar, y su carga iónica es la misma que la del Ho (Young, 2019). Aparte de ser un grupo de oligoelementos extremadamente coherentes, pueden ser utilizados como importantes trazadores geoquímicos para una amplia gama de investigaciones geológicas (Berger et al., 2014; Cole et al., 2014; Tsay et al., 2014; Craddock et al., 2010).

GRUPO

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

IA

IIA

IIIB

IVB

VB

VI

VII

VIII

VIIIB

IX

X

XI

XII

IIIB

IIIA

IVA

VA

VI

VII

VIII

PERIODO

1

2

3

4

5

6

7

1.008

6.94

22.990

39.098

85.468

132.91

223

4.0026

9.0122

24.305

24.305

40.078

87.62

137.33

178.49

232.04

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

238.03

Figura 2.7. Tabla periódica de elementos. Los recuadros rojos señalan los REE incluyendo al Sc y Y (extraído de Periodni s.f).

Las REE se clasifican en dos subgrupos en función del número atómico: tierras raras ligeras (LREEs: *Light Rare Earth Elements*; La, Ce, Pr, Nd, Pm, Sm y Eu); y tierras raras pesadas (HREEs: *Heavy Rare Earth Elements*; Gd, Tb, Dy, Ho, Er, Tm, Yb, y Lu) (Lopez et al., 2014). También existe un subgrupo conocido como tierras raras medianas (MREEs: *Middle Rare Earth Elements*) que es ocasionalmente aplicado de forma imprecisa a los elementos de alrededor de Pm y Ho (Henderson, 2013). Las REE poseen propiedades únicas que los hacen esenciales para la industria de alta tecnología por sus diversos usos en imanes permanentes de alta resistencia, láseres, convertidores catalíticos de automóviles, fibra óptica, superconductores y dispositivos electrónicos (Papangelakis y Moldoveanu, 2014; Zaremotlagh y Hezarkhani, 2016). La producción de minerales,

concentrados y metales REE está dominada por China, aunque existen varios yacimientos de alto valor económico de REE en otras localidades del mundo (Jaireth et al., 2014; Jha, 2014).

2.7.1. Comportamiento químico

Los REE son elementos químicos incompatibles, altamente electropositivos y con un grado de oxidación III (Balcerzak, 2000), que forman compuestos iónicos como óxidos, hidróxidos, haluros, carbonatos, fosfatos, silicatos, boratos y arsénicos, pero no sulfuros (Henderson, 1996; García-Bilbao, 2015). Se dividen en dos subgrupos, LREE y HREE, siendo el itrio parte de los HREE debido a sus propiedades químicas similares a los lantánidos (Voncken, 2016). El escandio, por su parte, tiene un comportamiento geoquímico más parecido a los elementos de transición ferromagnesianos (Felsche, 1971, 1973, 1978; Atwood, 2013), aunque a menudo se discute su inclusión en los REE debido a que en sistemas acuosos se comporta de manera similar a ellos (Beaudry y Geschneider, 1978; McLennan, 2012).

Los REE se encuentran juntos en la corteza terrestre debido a su carga trivalente y radios iónicos similares, y pueden ser fraccionados por procesos geológicos y termodinámicos (Zaremotlagh et al., 2016; Henderson, 1984, 1996; Jaireth et al., 2014; Rollinson, 1993). La distribución de los REEs es única en cada entorno geológico y puede ser influenciada por procesos como la meteorización y alteración (Baioumy et al., 2014; Cole et al., 2014; Ehya, 2012; Foley y Ayuso, 2013; Küpeli, 2010; Schacht et al., 2010; Shikazono et al., 2008). En sistemas magmáticos, la alta carga iónica de los REEs impide su equilibrio de carga y tienden a permanecer en el fundido coexistente durante el fraccionamiento cristalino (Van Gosen et al., 2017; Lipin y McKay, 2018). Factores como la temperatura, presión, composición del fluido, convección magmática y velocidad de sedimentación del mineral pueden influir en la partición de los REEs entre los minerales y el magma coexistente (Henderson, 1984; McKay, 1989; Kogarko et al., 2006).

La gran variedad de tipos y tamaños de los poliedros de coordinación de los cationes en los minerales que forman las rocas proporciona los medios para este fraccionamiento químico (Henderson, 2013). Otra de las características químicas de los REEs es que son litófilos, refractarios y tienden a ser elementos “dispersos”, debido a que se encuentran

presentes en cantidades traza en muchos minerales, pero raramente presentes en altas concentraciones y, con la excepción del Sc son incompatibles (Henderson, 2013; Lipin y McKay, 2018, Jochum et al., 1986; Atwood, 2013).

Los compuestos de REE tienden a ser relativamente insolubles en los fluidos acuosos y al igual que muchos metales traza, se presentan como cationes, que generalmente se vuelven menos solubles al aumentar el pH (Verplanck et al, 2004). Los coeficientes de partición fluido-roca tienden a ser muy bajos durante la mayoría de los procesos de interacción fluido-roca (Atwood, 2013). El comportamiento de los REEs en soluciones acuosas es relevante para la génesis de depósitos de minerales REE tanto en ambientes magmáticos como superficiales, ya que la estabilidad de los complejos acuosos de REE juega un papel importante en su movilización, transporte, fraccionamiento y deposición (Williams-Jones et al., 2012).

2.7.1.1. Óxidos de lantánidos

En la mayoría de los lantánidos, el estado de oxidación más estable es el 3+, por lo cual la gran mayoría, sino todos los óxidos de REE, se presentan como REE_2O_3 . A pesar de ello, algunos lantánidos pueden tener varias valencias en un mismo óxido por lo que se dan distintas fórmulas para expresar este fenómeno (Voncken, 2016). A temperatura ambiente, no todos los metales de REE son afectados de la misma manera por el aire. Los LREE tienden a oxidarse con más rapidez, especialmente el europio, el lantano y el neodimio. La oxidación se multiplica por diez cuando la humedad relativa aumenta del 1 al 75% (Gupta y Krishnamurthy 2005).

2.7.1.2. Contracción de los lantánidos

Los lantánidos (del lantano al lutecio) tienen una propiedad única: con cada aumento del número atómico en el grupo, se añade un electrón a un subesqueleto interno incompleto (4 f) en lugar de a un nivel externo. Esta propiedad de llenar selectivamente los niveles internos da lugar a una disminución progresiva e importante del radio iónico de los iones lantánidos trivalentes de La^{3+} a Lu^{3+} (Figura 2.8), lo cual es denominado como “contracción de los lantánidos” (Van Gosen et al., 2017; Voncken, 2016).

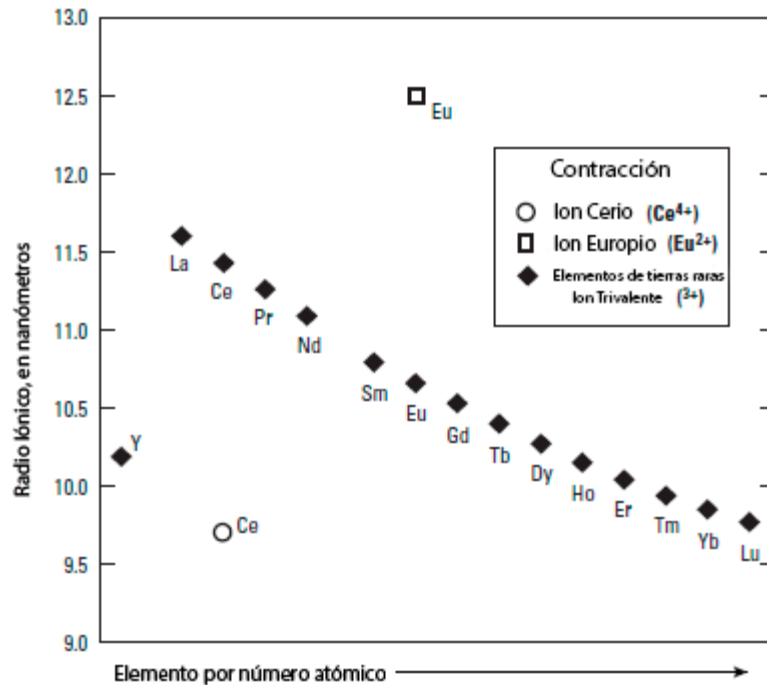


Figura 2.8. Gráfico que muestra los radios (en nanómetros) de los iones trivalentes ($3+$) de los elementos de tierras raras (REE), así como del cerio en el estado de valencia $+4$ (Ce^{4+}) y del europio en el estado de valencia $+2$ (Eu^{2+}). El gráfico ilustra la disminución sistemática del radio de los iones REE trivalentes con el aumento del número atómico. Los datos representados corresponden al número de coordinación VIII. Los datos proceden de Shannon (1976). El prometio (Pm) no está incluido en la lista porque es extremadamente raro en la naturaleza (Extraído de Van Gosen et al., 2017).

Tabla 2.1. Concentración de REE en la corteza vs concentración en Proyectos de Exploración y Explotación, en ppm (modificado de Long et al 2011 y López, 2017).

Abundancia de REE en la Corteza Vs Proyectos de Exploración y Explotación															
Abundancia de REE en la Corteza	Autor	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
	Wedephol (1995)	30	60	6,7	27	5,3	1,3	4	0,65	3,8	0,8	2,1	0,3	2	0,35
	Lide (1997)	39	66,5	9,2	41,5	7,05	2	6,2	1,2	5,2	1,3	3,5	0,52	3,2	0,8
	McGill (1997)	5 a 18	20 a 46	3,5 a 5,5	12 a 24	4,5 a 7	0,14 a 1,1	4,5 a 6,4	0,7 a 1	4,5 a 7,5	0,7 a 1,2	2,5 a 6,5	0,2 a 1	2,7 a 8	0,8 a 1,7
Proyectos de exploración y explotación	Tipo de depósito														
Bear Lodge (EE.UU.)	Hierro-REE	6010,92	9475,99	1092,56	4158,15	681,29	155,45	364,39	34,01	95,84	8,73	17,49	0,00	8,78	0,00
Araxá (Brasil)	Carbonatita	10061,62	16924,91	1580,91	4998,36	543,31	120,91	251,60	25,51	104,56	17,46	26,24	0,00	17,56	0,00
Browns Range (Australia)	Placer	136,43	309,35	41,39	205,76	112,11	25,91	294,98	68,02	461,80	96,03	288,59	43,78	237,11	35,18
Cummins Range (Canadá)	Carbonatita	3990,54	6626,68	695,27	2340,57	284,59	51,82	173,52	0,00	78,42	0,00	8,75	0,00	0,00	0,00
Glenover (Sudáfrica)	Vetas de fosfatos	3061,12	7782,69	1009,79	4055,27	646,79	164,09	373,06	34,01	139,41	17,46	34,98	0,00	17,56	0,00
Elliot Lake Teasdale (Canadá)	Depósito asociado con Uranio	409,29	716,40	74,49	240,06	43,12	0,00	26,03	0,00	17,43	0,00	8,75	0,00	8,78	0,00
Mount Weld CLD (Australia)	Carbonatita	19807,76	37651,61	4155,06	15123,68	2043,86	449,08	919,65	76,52	209,12	26,19	52,47	8,76	26,35	0,00
Mount Weld Duncan (Australia)	Carbonatita	10249,21	15500,25	1903,71	7416,09	1181,47	319,54	832,89	110,53	531,50	78,57	174,90	17,51	79,04	8,79
Mountain Pass (EE.UU.)	Carbonatita	18588,42	26246,23	2334,12	6755,93	457,07	60,45	112,79	34,01	26,14	8,73	8,75	8,76	8,78	8,79
Songwe Hill (Malawi)	Adsorción Ionica	3103,76	5364,85	587,67	2074,79	301,84	69,09	173,52	17,00	104,56	17,46	34,98	8,76	26,35	0,00
Strange Lake Enriched (Canadá)	Complejos alcalinos	1279,02	2930,72	322,80	1200,29	310,46	17,27	338,36	76,52	575,07	130,95	428,52	70,05	447,88	61,56

2.7.2. Depósitos

Los depósitos de REE se presentan en rocas sedimentarias, metamórficas e ígneas, siendo las últimas las de mayor importancia y asociadas normalmente en variedades poco comunes de rocas ígneas, tales como rocas alcalinas y carbonatitas (López, 2017; Long et al., 2012; Jones et al., 1995; Haxel, 2005, Haxel et al., 2005). En la Tabla 2.1 se puede observar que la concentración de REE en ppm en la corteza varía según algunos autores, pero presentan rangos similares. En cambio, los rangos de leyes en proyectos de exploración y explotación suelen ser exponencialmente mayores en la mayoría de los casos. El Lu y el Tm son los elementos que presentan una menor o ausencia de concentración de valor económico. El mayor depósito de minerales de REE en el mundo es Bayan Obo, en la Región Autónoma de Nei Mongol, China (Hitzman, 1992; Atwood, 2013; Fan et al., 2016; Wu et al., 1995). Este yacimiento es una masa compleja e irregular de depósitos de hierro y carbonatita mezclados (Schulz et al., 2018). Diversos autores han descrito su geología y mineralogía (Zhongxin et al., 1992; Xu et al., 2010; Yang et al., 2010, 2011; Yang et al., 2013; Kynicky et al., 2012; Verplanck et al, 2014; Goodenough, 2017). Se cree que la gran variedad de minerales (28) de REE presentes, puede ser producto de metamorfismo posterior a la mineralización (Qui, 1990; Zeng, 1990).

Este apartado no pretende abarcar todos los yacimientos geológicos subeconómicos en los que se han encontrado REEs, sino que se enfocará en los depósitos de carbonatitas, peralcalinos, magnetita-hematita magmáticos y depósitos IOCG. Van Gosen et al (2017) señalan que los depósitos de REE potencialmente económicos se encuentran principalmente en los siguientes entornos geológicos:

- Carbonatitas
- Sistemas ígneos peralcalinos
- Cuerpos magnetita-hematita magmáticos
- Depósitos de óxido de hierro-cobre-oro (IOCG)
- Acumulaciones de xenotiempo-monacita en gneis máficos
- Depósitos de arcilla de absorción iónica
- Depósitos aluviales de monacita
- Xenotiempo

2.7.2.1. Formación de REE en magmas de carbonatita

Desde la década de 1960 las carbonatitas han sido la principal fuente mundial de LREEs (U.S. Geological Survey, 2015). Estas son rocas derivadas de magmas ricos en carbonatos con más del 50% modal de minerales de carbonato, habitualmente calcita y dolomita, las intrusiones de carbonatita se presentan en una variedad de formas y composiciones de carbonatos, y en general, presentan las mayores concentraciones de REE de todas las rocas ígneas (Chakhmouradian y Zaitsev, 2012; Voncken, 2016). Los conjuntos de minerales de REE en las carbonatitas pueden ser clasificadas en tres grupos según el proceso de formación del mineral (Mariano, 1989; Chakhmouradian y Zaitsev, 2012):

1. Mineralización magmática primaria en carbonatita (Mountain Pass, California).
2. Mineralización hidrotermal-metasomática (Bayan Obo, China).
3. Mineralización supergénica (Mount Weld en Australia Occidental, Australia, y Araxá en el Estado de Minas Gerais, Brasil).

En los yacimientos de carbonatita, los minerales principales de REE son los fluorocarbonatos (bastnaesita, parisita, sinquisita), los carbonatos hidratados (ancilita) y los fosfatos (monacita) (Mariano 1989; Mariano y Mariano Jr. 2012). La bastnaesita es la fuente más significativa de REEs en estos depósitos (Van Gosen et al., 2017; Gupta y Krishnamurthy 2005).

2.7.2.2. Formación de REE en magmas alcalinos

Los REE tienen una fuerte asociación genética con los procesos ígneos alcalinos, específicamente el magmatismo peralcalino. Las rocas ígneas peralcalinas normalmente se asocian con enriquecimiento de REE más que la mayoría de las demás rocas ígneas (U.S. Geological Survey, 2015). La gran mayoría de los yacimientos peralcalinos de REE exhiben una evolución petrogenética compleja, que puede dar lugar a una mineralogía variable y a texturas intrincadas. En general, la mineralización de REE en las intrusiones peralcalinas están más enriquecidas de HREEs a diferencia de los depósitos alojados en carbonatita, por lo mismo han sido objeto de exploraciones recientes (Van Gosen et al., 2017). Sin embargo, los procesos de transporte y enriquecimiento de REE en los sistemas peralcalinos siguen siendo poco conocidos (Chakhmouradian y Zaitsev, 2012).

2.7.2.3. Cuerpos de magnetita-hematita magmáticos

Los cuerpos de magnetita-hematita magmáticos pueden contener minerales portadores de REE, con el potencial de recuperar los REE como subproducto durante la extracción de hierro (e.g. distritos mineros de Pea Ridge y Mineville) (Long et al., 2012). Los minerales portadores de REE identificados en estos depósitos son principalmente monacita, xenotima y cantidades menores de bastnaesita y britholita (Seeger et al., 2001). Los depósitos de hierro se caracterizan por magnetita, martita y apatita, con augita accesorio, hornblenda, albita, cuarzo, pirita y turmalina (McKeown y Klemic, 1956). En Chile los depósitos Cerro Carmen y Sierra Áspera presentan el potencial más interesante de REE.

2.7.2.4. Depósitos de óxido de hierro-cobre-oro (IOCG)

Los yacimientos de IOCG están estrechamente relacionados con los yacimientos de hierro, mencionados con anterioridad, y se podría argüir que los yacimientos de Pea Ridge y Mineville representan variaciones dentro del grupo de IOCG (U.S. Geological Survey, 2015). Los REE están presentes en concentraciones anómalas en todos los yacimientos de óxidos de hierro en los que se dispone de datos analíticos, siendo el yacimiento de Bayan Obo el más reconocido (Argall, 1980).

Los depósitos de IOCG pueden albergar mineralización de REE, que se produce tanto en óxidos de hierro como en pequeñas carbonatitas dentro del complejo intrusivo. A pesar de que se han identificado recursos de REE de tamaño considerable en Olympic Dam, aún no se ha recuperado REEs (Van Gosen et al., 2017). Las brechas de Olympic Dam contienen una media de REE de aproximadamente 5000 ppm, donde las mayores concentraciones se localizan en las brechas más hematíticas del núcleo del yacimiento (Oreskes y Einaudi, 1990).

Además de los depósitos IOCG, en los depósitos de óxido de hierro-apatito (IOA) los REEs son un importante subproducto de estos depósitos (Skirrow, 2014; Simandl, 2014). El enriquecimiento en REE está intensamente asociado a la formación de minerales de fosfato en muchos depósitos de IOA. Sin embargo, la bastnaesita, el praseodimio y la allanita son tan significativos en algunos lugares (Oreskes y Einaudi, 1990). Algunas investigaciones interpretan que la lixiviación de REE post-deposicional podría ocurrir en el apatito en la que se pueden ver las inclusiones de monacita y xenotima (Bonyadi et al., 2011; Stosch et al.,

2011; Torab y Lehmann, 2007). El apatito está presente en la mayoría de los depósitos de óxido de hierro del norte y centro de Suecia (Frietsch, 1979), y las leyes de REE son comparables a los de Kiruna (Parak, 1973). La datación U-Pb de las inclusiones de monacita en el apatito demuestra que la redistribución de REE en el apatito puede ocurrir con frecuencia a lo largo del proceso hidrotermal varios millones de años después de la formación de los depósitos de IOA (Stosch et al., 2011).

2.7.3. Mineralogía

Los minerales que presentan REEs son numerosos, diversos y a menudo de composición compleja. Algunos autores enumeran 245 minerales individuales portadores de REE, incluyen silicatos (85); óxidos (59); carbonatos, fluorocarbonatos e hidroxycarbonatos (42); fosfatos (26); arseniatos, sulfatos y vanadatos (19); haluros (6); carbonatos de uranilo y silicatos de uranilo (5); y boratos (3) (Jones et al., 1996; Möller et al, 2013).

Las principales menas de REE, corresponden a los minerales de monacita ((REE, Th, Y)PO₄) y xenotima, ambos fosfatos, bastnasita (REE(CO₃F)-carbonato) y allanita (Ca(Ce,La,Y)(Al₂Fe²⁺)-silicato). La Tabla 2.2 presenta los minerales de REE con un alto potencial económico (López, 2017). Las fuentes más significativas desde el punto de vista se clasifican de la siguiente forma (Kanazawa y Kamitani, 2006; López, 2017; Avendaño, 2017):

- **Monacita:** La monacita es mineral de fosfato de fórmula generalizada CePO₄, además de Ce, también hay otros REE en la monacita, siendo principalmente LREEs (La, Ce, Pr, Nd y Sm) (Voncken, 2016). Corresponde a la segunda mena más importante de LREEs con reservas en Australia, Brasil, China, India, Sri Lanka y Estados Unidos (U.S. Geological Survey, 2015). Los yacimientos que contienen monacita suelen contener torio y/o uranio, elemento radioactivo, pero las cantidades son generalmente demasiado bajas para extraerlas como subproducto valioso, sumado a los daños medioambientales que puedan provocar, producto de ello muchos países lo descartan como fuente de REE (Humphries, 2011; Gupta y Krishnamurthy 2005).
- **Bastnaesita:** Descrita por primera vez por Wilhelm Hisinger en 1838. Es un mineral de fluorocarbonato (Ce(CO₃F) bastante extendido, cuyos depósitos están presentes en China y Estados Unidos, y presentan las mayores reservas de LREEs en el mundo

(U.S. Geological Survey , 2015). De los HREE, solo contiene Y. Se encuentra en una gran variedad de rocas ígneas, como las carbonatitas, los depósitos de vetas, las rocas metamórficas de depósitos de vetas, rocas metamórficas de contacto y pegmatitas (Gupta y Krishnamurthy 2005; Voncken, 2016). Los fuertes enriquecimientos en LREEs ocurren en depósitos de carbonatita en los que el mineral de mena predominante es la bastnaesita (Mariano, 1989).

- **Xenotima:** La xenotima es un fosfato $(\text{REE})\text{PO}_4$, descrito por primera vez por Berzelius en Flekkefiord Vest-Agder, Noruega (Berzelius 1824, 1825; Voncken, 2016). Es rico en Y con un 25 - 60% en peso, y suele contener cantidades apreciables de HREE (Y, Tb, By, Ho, Er, Tm, Yb y Lu) convirtiéndolo en un mineral de especial interés (Dutta et al., 2016; Atwood, 2013). La xenotima es un mineral accesorio en pegmatitas y otras rocas ígneas no básicas, es muy similar a la monacita, además se caracteriza por una alta gravedad específica (Gratz y Heinrich, 1997).
- **Eudialita:** La Eudialita es un ciclosilicato con la fórmula general $\text{Na}_4(\text{Ca}, \text{Ce})_2(\text{Fe}^{2+}, \text{Mn}^{2+}) \text{ZrSi}_8\text{O}_{22}(\text{OH}, \text{Cl})_2$, descrita por primera vez por Klaproth (1810). (Anthony et al., 2001; Mindat.org, 2022). La eudialita ígnea se presenta asociada en pegmatitas (Deer et al., 1986a).
- **Euxenita:** mineral de fórmula $[(Y, \text{Ce}, \text{U}, \text{Th})(\text{Nb}, \text{Ta}, \text{Ti})_2\text{O}_6]$. Habitualmente se encuentra como tantalonibates (minerales donde Ta y Nb forman el compuesto) de titanio, tierras raras, torio y uranio (Avendaño, 2017).
- **Allanita:** Es un mineral del grupo de las epidotas con fórmula química $[(\text{Ce}, \text{La}, \text{Y})(\text{Al}_2\text{Fe}^{2+})(\text{O}/\text{OH}/\text{SiO}_4) \cdot \text{Si}_2\text{O}_7]$ (Avendaño, 2017). Está presente en tres formas minerales, las cuales contienen cerio, lantano e itrio. Se encuentra en ambientes ígneos, metamórficos e hidrotermales (Weber y Reisman, 2012; Deer et al., 1986b).

Tabla 2.2. Minerales que presentan REE con un alto potencial económico. Los nombres están escritos en inglés, dado que no todos cuentan con una traducción literal al español (extraído de López, 2017).

Mineral	Formula*	Óxidos de tierras raras
		wt %† ‡
Aeschynita	$(\text{Ln}, \text{Ca}, \text{Fe}, \text{Th})(\text{Ti}, \text{Nb})_2(\text{O}, \text{OH})_6$	36
Allanita (Ortita)	$(\text{Ca}, \text{Ln})_2(\text{Al}, \text{Fe})_3(\text{SiO}_4)_3(\text{OH})$	30
Anatasa	TiO_2	3
Ancylita	$\text{SrLn}(\text{CO}_3)_2(\text{OH}) \cdot \text{H}_2\text{O}$	46
Apatita	$\text{Ca}_5(\text{PO}_4)_3(\text{F}, \text{Cl}, \text{OH})$	19
Bastnasita	LnCO_3F	76
Brannerita	$(\text{U}, \text{Ca}, \text{Ln})(\text{Ti}, \text{Fe})_2\text{O}_6$	6
Britholita	$(\text{Ln}, \text{Ca})_5(\text{SiO}_4, \text{PO}_4)_3(\text{OH}, \text{F})$	62
Cerianita	$(\text{Ce}, \text{Th})\text{O}_2$	81§
Cheralita	$(\text{Ln}, \text{Ca}, \text{Th})(\text{P}, \text{Si})\text{O}_4$	5
Churchita	$\text{YPO}_4 \cdot 2\text{H}_2\text{O}$	44†
Eudialyte	$\text{Na}_{15}\text{Ca}_6(\text{Fe}, \text{Mn})_3\text{Zr}_3(\text{Si}, \text{Nb})\text{Si}_{25}\text{O}_{73}(\text{OH}, \text{Cl}, \text{H}_2\text{O})_5$	10
Euxenita	$(\text{Ln}, \text{Ca}, \text{U}, \text{Th})(\text{Nb}, \text{Ta}, \text{Ti})_2\text{O}_6$	<40§
Fergusonita	$\text{Ln}(\text{Nb}, \text{Ti})\text{O}_4$	47
Florencia	$\text{LnAl}_3(\text{PO}_4)_2(\text{OH})_6$	32§
Gadolinita	$\text{LnFeBe}_2\text{Si}_2\text{O}_{10}$	52
Huanghoita	$\text{BaLn}(\text{CO}_3)_2\text{F}$	38
Hydroxylbastnasita	$\text{LnCO}_3(\text{OH}, \text{F})$	75
Kainosita	$\text{Ca}_2(\text{Y}, \text{Ln})_2\text{Si}_4\text{O}_{12}\text{CO}_3 \cdot \text{H}_2\text{O}$	38
Loparita	$(\text{Ln}, \text{Na}, \text{Ca})(\text{Ti}, \text{Nb})\text{O}_3$	36
Monazita	$(\text{Ln}, \text{Th})\text{PO}_4$	71
Mosandrita	$(\text{Ca}, \text{Na}, \text{Ln})_{12}(\text{Ti}, \text{Zr})_2\text{Si}_7\text{O}_{31}\text{H}_6\text{F}_4$	<65§
Parisita	$\text{CaLn}_2(\text{CO}_3)_3\text{F}_2$	64
Samarskita	$(\text{Ln}, \text{U}, \text{Fe})_3(\text{Nb}, \text{Ta}, \text{Ti})_5\text{O}_{16}$	12
Synchisita	$\text{CaLn}(\text{CO}_3)_2\text{F}$	51
Thalenita	$\text{Y}_3\text{Si}_3\text{O}_{10}(\text{OH})$	63§
Xenotima	YPO_4	61§
Yttrotantalita	$(\text{Y}, \text{U}, \text{Fe})(\text{Ta}, \text{Nb})\text{O}_4$	<24§

* Formulas del mineral: Mandarin (1999), Ln= Elementos lantánidos

† Contenidos de Óxidos de tierras raras: Frondel, 1958; Overstreet, 1967; Anon, 1980; Kapustin, 1980; Mazzi y Munno 1983; Mariano 1989.

‡ Cuando más de un análisis está disponible, se ha tomado el análisis con mayor contenido de óxidos de REE (por ejemplo, en los óxidos de REE para la monacita en Mountain Pass se reportó la carbonatita; la monacita de pegmatitas y en rocas metamórficas, generalmente tiene menor concentración en los óxidos de REE).

§Cálculo estequiométrico del contenido de óxidos de REE.

2.7.4. Aplicaciones

Como se ha mencionado anteriormente, los REE juegan un papel estratégico para el desarrollo de los países que poseen estos recursos, junto al rol sustancial que tendrán en nuevas tecnologías, debido que constituyen materiales imprescindibles e insustituibles en la actual industria (López, 2017; Goonan, 2011; Long, 2011). Sus usos y aplicaciones en la industria varían entre productos químicos, electrónica, cerámicas, vidrios, iluminación, imanes, técnicas medias, etc (Castor y Hendrick, 2006; Massari y Ruberti, 2013; Stegen, 2015). La industria del vidrio es la principal consumidora de materias primas de REE, primariamente para el pulido del vidrio y como aditivos para proveer color y propiedades ópticas especiales al vidrio (Goonan, 2011; Van Gosen, 2017). La Tabla 2.3 muestra algunos de sus usos. La demanda de estos elementos se ha incrementado drásticamente dado a sus amplios usos y versatilidad (López, 2017). Según De la Hoz (2012), dado a los motivos expuestos previamente, estos elementos son catalogados como estratégicos para el desarrollo de las sociedades modernas.

Tabla 2.3. Principales usos de REE. Extraída de Cortés y del Valle Giles (2014).

Tierras raras ligeras	Principales usos
Escandio	Luces de estadios deportivos
Lantano	Motores híbridos, baterías híbridas y aleaciones metálicas
Cerio	Catalizadores de automóviles, refinерías de petróleo, aleaciones metálicas, motores híbridos y baterías híbridas
Praseodimio	Imanes, motores híbridos, baterías híbridas, discos duros de computadora, teléfonos móviles, cámaras, reflectores, piezas de aviones
Prometio	Unidades portátiles de rayos X
Neodimio	Imanes de alta fuerza, catalizadores de automóviles, refinерías de petróleo, discos duros de computadora y laptops, teléfonos móviles, auriculares, cámaras, motores híbridos, baterías híbridas
Samario	Imanes
Europio	Color rojo para pantallas de televisión, de computadora y fibra óptica
Tierras raras pesadas	Principales usos
Gadolinio	Imanes
Disprosio	Imanes permanentes, motores híbridos, baterías híbridas, discos duros de ordenador, teléfonos móviles y cámaras
Terbio	Fósforos, imanes permanentes, discos duros de computadora, teléfonos móviles y cámaras
Erbio	Fósforos
Itrio	Color rojo, lámparas fluorescentes, cerámica, aleaciones metálicas, fibra óptica
Holmio	Coloración de vidrio, láser
Tulio	Instrumentos médicos como rayos X
Lutecio	Catalizadores en la refinación de petróleo
Yterbio	Láser y aleaciones de acero

2.7.5. Mercado Mundial

La explotación de REE comenzó en el siglo XIX y tuvo un impacto importante en la industria (López, 2017). China es el principal productor y controla alrededor del 90% de la producción mundial (Papangelakis y Moldoveanu, 2014; U.S. Geological Survey, 2015). Australia lideraba en depósitos de placer de minerales pesados con tierras raras como subproducto (López, 2017), pero desde la década de 1980 China emergió como un importante productor y tomó el control en un corto período de tiempo (Gabás, 2020). En 1990, la Conferencia Nacional de la Industria del Metal No Ferroso China declaró a las REE como recursos estratégicos y limitó su exportación, controlando su mercado mundial (Butler, 2014). Estados Unidos y Australia sufrieron una caída en la producción en contraste con China (Figura 2.9) (Cortés y del Valle Giles, 2014; López, 2017; Hurst, 2010).

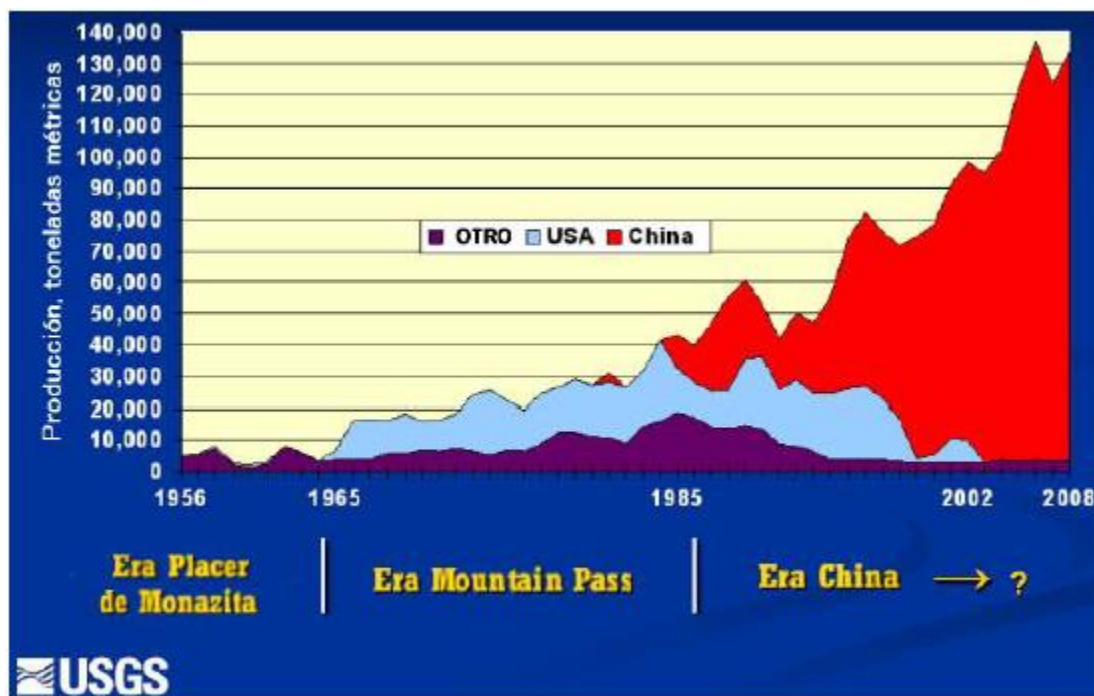


Figura 2.9. Producción total de óxidos de REE (extraído de López, 2017, modificado de U.S. Geological Survey, 2005).

En los años 2000, China se convirtió en el máximo productor y exportador de REE, monopolizando el mercado hasta la fecha (Avendaño, 2017). En 2014, la producción de China representó el 85% de la producción mundial (U.S. Geological Survey, 2015). Los recursos de REE se han descubierto en 21 provincias de China (U.S. Geological Survey, 2015), y los yacimientos de arcilla de adsorción iónica tienen concentraciones entre 300 ppm

y 2.000 ppm (Bao y Zhao, 2008). Esto ha generado preocupación en los países industrializados por la importancia estratégica de estos elementos, lo que ha llevado a un mayor énfasis en la exploración y extracción de REE fuera de China (Mancheri, 2015). Se espera que la demanda de REE crezca notablemente debido al auge de nuevas tecnologías, como los autos eléctricos y las energías renovables. Además, se pronostica que la participación de China en la producción mundial se verá afectada debido a nuevas iniciativas en otros países (Roskill, 2018). La Figura 2.10 muestra la diferencia entre la producción mundial vs China entre los años 2000 – 2015, donde se aprecia claramente la influencia de China en la oferta global (Dutta et al., 2016).

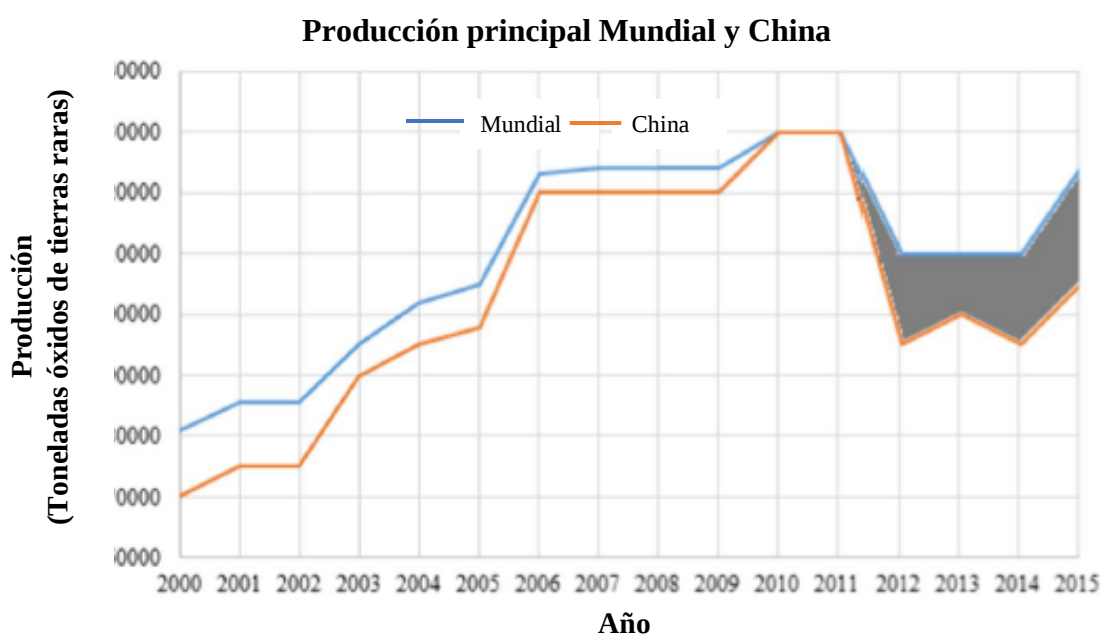


Figura 2.10. Tendencia histórica de la producción de REE (en toneladas) en el mundo y en China (según el U.S. Geological Survey). La zona sombreada indica la disminución de la cuota de China en la producción mundial de REE, que pasó de más del 95% antes de 2012 al 85% en 2015 (modificado de Dutta et al., 2016).

2.7.6. Tierras raras en Chile

Si bien Chile no presenta yacimientos con grandes concentraciones de REE, existe un gran potencial a explorar y se ha iniciado su investigación (López et al., 2014). Según COCHILCO (2016), en Chile las reservas conocidas de REE se ubican principalmente en las regiones de Atacama y Biobío. Debido a la caída de las leyes de cobre y al incremento de los costos asociados a su explotación, han surgido distintas iniciativas que buscan diversificar la

exploración y explotación de nuevos recursos los cuales puedan entregar mayores beneficios económicos (MCH, 2015).

En 2011, Sernageomin (Servicio Nacional de Geología y Minería) inició el desarrollo de un mapa geoquímico para promover la exploración minera y un desarrollo sustentable en Chile (Prieto, 2014). Actualmente, se ha cubierto una superficie de 108.400 km² en 8 hojas con un promedio de 677 puntos de muestreo cada una (SERNAGEOMIN, 2020). El estudio ha indicado zonas de interés en el norte del país con concentraciones anómalas relativamente altas de REE, especialmente en yacimientos de tipo IOA y IOCG (SERNAGEOMIN, 2014; Prieto, 2014). Además, un proyecto CORFO llevado a cabo por la Cámara Chileno Alemana de Comercio, identificó la presencia de cerio, lantano y neodimio en relaves de la región de Atacama (CORFO, 2013).

Debido a la alta demanda y al interés de reducir la dependencia de China, hoy en día existen distintas iniciativas para la extracción de concentraciones económicamente rentable (Hatch y Lifton, 2013). Es por ello por lo que en el año 2011 el conglomerado Minería Activa levantó un Fondo de Inversión Privado (FIP) de lantánidos, con el objetivo de desarrollar la minería de REE en Chile (Biolantánidos; Avendaño, 2017). En este marco se llevó a cabo el proyecto Aclara (ex-Biolantánidos) que busca entrar en el mercado de REE (Vega, 2021; MCH, 2015). El proyecto se encuentra ubicado en el Cabrito, Penco, Región del Bío Bío y esperan estar en producción a principios de 2023 (Munita, 2021). Con este proyecto Chile podría convertirse en uno de los escasos proveedores de REE (Vega, 2021).

Se estima que en Chile existe un gran potencial de REE en relaves (750), los cuales aún no han sido extraídos. La viabilidad económica para procesar los relaves está relacionada con su contenido de cobre (MCH, 2020; BNamericas, 2020). El tranque Coemin en la región de Atacama es el relave más rico en REE, con una ley de cobre baja (0,2%), donde el La (176,7 ppm) y el Ce (288,1 ppm) presentan las mayores concentraciones seguido por el Nd (56,2 ppm) (Chávez, 2017), pero otros relaves como el de Planta Vallenar y El Salado también contienen valores significativos de REE (370 y 350 ppm en promedio entre La y Ce, respectivamente) y leyes interesantes de cobre (4,1 y 1,2% de cobre) (Riquelme, 2020). No se ha encontrado evidencia de leyes de REE en los concentrados de la Mina Candelaria y se

cree que estos elementos deberían estar presentes en zonas de mayor profundidad (Prieto, 2014).

2.8. Yacimientos tipo IOCG

Los depósitos tipo IOCG (*Iron Oxide-Copper-Gold*) fueron reconocidos como una clase distinta de yacimientos tras el descubrimiento del gigantesco yacimiento de Cu-U-Au-Ag-REE de Olympic Dam en el sur de Australia en 1975 por parte de Western Mining Corporation (Hitzman et al., 1992). Olympic Dam contiene una de las cinco mayores reservas de cobre del mundo y la mayor reserva de uranio extraído (Mudd y Jowitt, 2018). Producto del descubrimiento de nuevos depósitos de este tipo, ha aumentado su atención considerablemente a nivel mundial, tanto por parte de las instituciones académicas como de empresas de exploración (Williams et al., 2005). El modelo IOCG, más que un modelo geológico estrictamente definido, alude a la asociación de mineralización de cobre y oro con óxidos de hierro, magnetita o hematita (Oyarzún, 2007).

El conocimiento de los depósitos de IOCG se ha profundizado gradualmente en las últimas dos décadas (Williams et al., 2005; Groves et al., 2010). Estos depósitos son una de las principales fuentes de Cu y Au del mundo, y están caracterizados por abundante mineralización de óxidos de Fe, en su mayoría magnetita y/o hematita especular, acompañada de minerales de mena como calcopirita $\pm$ bornita y escasa presencia de sulfuros de Fe, pobre en Ti, y a diferencia de otros tipos de depósitos, no parecen tener una relación espacial y genética clara entre la mineralización e intrusivos (Porter, 2000; Ray y Lefebure, 2000; Oberreuter, 2017). Algunos depósitos de IOCG también contienen importantes recursos de Ag, Bi, Mo y "minerales críticos" como REE y el cobalto (Skirrow, 2019).

Los principales yacimientos de cobre-oro de estilo IOCG están asociados temporalmente a granitoides oxidados y potásicos similares a los vinculados a los principales pórfidos de cobre-oro, además de existir una asociación con los principales sistemas de fallas que podrían haber actuado como sistemas de cizallamiento frágil-dúctil de tipo transpresional a transtensional en el momento de la intrusión del granitoide y la mineralización (Pollard, 2000). Comprenden muchos sistemas de mineralización diferentes en el tipo, ocurrencia y alteración, debido a las condiciones variables de presión y temperatura reflejándose en un amplio espectro de depósitos; encontrándose en todos los continentes, generalmente en rocas

con edades desde el Proterozoico temprano hasta el Plioceno (Hitzman et al., 1992, 2000; Barton y Johnson, 1996). Además, cuentan con una variada morfología, que va desde cuerpos estratoligados hasta *stockwork* en zonas de brechas irregulares (Barra et al., 2017; Makshev y Zentilli, 2002).

Tras años de estudios, se ha creado una fuerte base descriptiva de los yacimientos de IOCG, pero a pesar de ello, aún se presentan problemas sin resolver sobre la definición, clasificación, génesis y objetivos de exploración de esta amplia y diversa familia de depósitos (Skirrow, 2019). Muchos de los problemas expuestos nacen de la limitada comprensión de las similitudes y diferencias fundamentales entre la amplísima gama de depósitos de óxidos de hierro de cobre y oro (Groves et al., 2010; Porter, 2010).

2.8.1 Génesis

La proveniencia de los fluidos mineralizadores de estos sistemas ha sido tema de discusión; se ha postulado que provienen de salmueras formadas en cuencas (Barton y Johnson, 2000), en cambio otros autores han asociado la mineralización a la actividad hidrotermal producto del magmatismo (Sillitoe, 2003) (ver esquema en Figura 2.11). Algunos autores favorecen los modelos no magmáticos, que pueden subdividirse en dos categorías: aquellos en que los fluidos provendrían de salmueras principalmente superficiales, de cuencas poco profundas y/o evaporitas, movilizados termalmente para lixiviar y re-depositar metales y azufre; y aquellos en el que los fluidos han evolucionado en entornos metamórficos de grado bajo o medio (Battles y Barton 1995; Barton y Johnson 1996, 2000; Barton et al., 1998; Ulrich y Clark, 1999; Williams et al., 2005; Chen et al., 2011). En cambio, otros autores han asociado la mineralización a la acción de actividad hidrotermal producto del magmatismo donde se liberan salmueras oxidadas, pobres en azufre y metalíferas coetáneas, las cuales son exsuestras desde un magma en cristalización, y que luego depositan metales en las rocas de caja adyacentes (Gow et al., 1994; Rotherham et al., 1998; Williams, 2010; Williams y Skirrow, 2000; Skirrow, 2007).

La mineralización incluye una fase inicial compuesta por óxidos de hierro, principalmente magnetita y hematita especular, seguido por una fase rica en sulfuros de Cu y Fe, como piritita, calcopiritita y mushketovita (Oberreuter, 2017). Además, se ha planteado que los controles composicionales y petro-tectónicos serían relativamente poco importantes, mientras que los

sistemas que favorezcan la penetración de salmueras externas como lo son sistemas de bombeo tectónico, controles paleogeográficos y paleoclimáticos serían críticos (Barton y Johnson, 1996, 2000, 2004; Barton, 2014).

Ha sido cuestionado la importancia de las secuencias evaporíticas como requerimiento para la formación de depósitos IOCG, debido a la ausencia de éstas en algunos distritos, resultando fuentes más bien magmáticas para los fluidos formadores de mena, así también como para los metales (Pollard, 2000, 2006). Se han propuesto modelos en función de la temperatura y la profundidad, variando desde las más altas temperaturas y profundidades, en conjunto de fugacidades de oxígeno relativamente bajas (e.g. Ernest Henry), hasta los depósitos formados en niveles más superficiales de la corteza por medio de fluidos con temperaturas menores y más oxidados (e.g. Olympic Dam). Se ha sugerido que el *mixing* o mezcla de fluidos podría ser el causante de la variedad de depósitos IOCG (Figura 2.12.) (Mark et al., 2000; Barton y Johnson, 2004).

2.8.2 Alteración

Estos yacimientos son bien conocidos por sus amplios halos de alteración en especial la alteración alcalina (cálcica), y también su marcada zonificación, como se observa en los sistemas de IOCG del yacimiento de Olympic Dam (Ehrig et al., 2012; Mauger et al., 2016; Dmitrijeva et al., 2019a) o Ernest Henry, en Queensland (Mark et al., 2006). Las diferentes alteraciones hidrotermales han sido ampliamente descritas por diversos autores (Barton y Johnson, 2000; Williams et al., 2005; Sillitoe, 2003).

Las rocas huésped en las cercanías de los depósitos presentan una intensa alteración hidrotermal. Comúnmente se observa alteración calco-sódica (Ca-Na), y suele presentar una asociación mineral de albita, escapolita, actinolita, clorita y epidota, con mineralización de magnetita (Dmitrijeva et al., 2019a; Zandiyyeh et al., 2016; Oberreuter, 2017). La alteración del Na y del Ca-Na, cuando están presentes, tienden a desarrollarse a mayor profundidad o a mayor distancia de la mena, es más extensa y suele ser anterior a la alteración del K-Fe y a la mineralización (Dmitrijeva et al., 2019b). Los carbonatos son comúnmente abundantes, particularmente en asociación con sulfuros portadores de Cu (Williams et al., 2005).

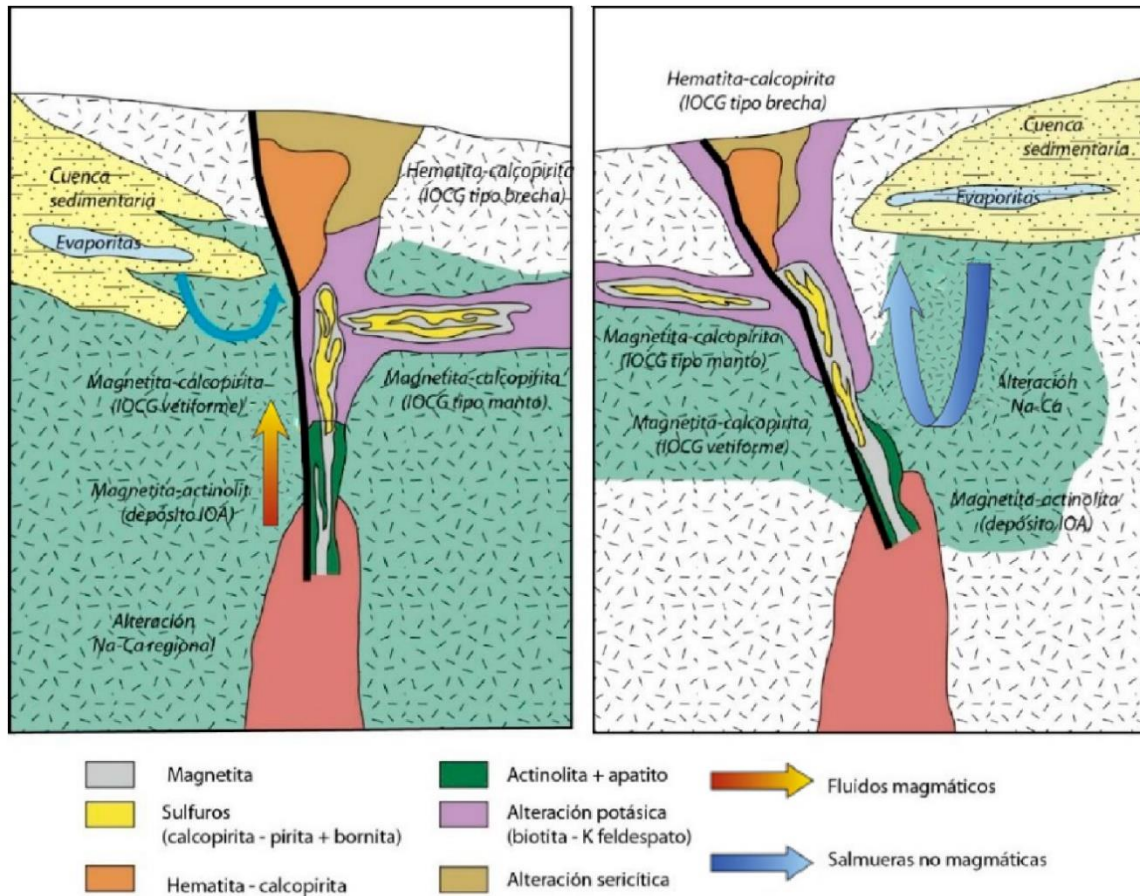


Figura 2.11. Representación esquemática de los dos modelos propuestos para el origen de los depósitos de IOCG. A) Modelo magmático-hidrotermal: los fluidos y los metales son de origen magmático con una contribución menor de rocas sedimentarias y/o volcánicas. B) Los fluidos proceden de cuencas sedimentarias y son calentados por una fuente ígnea. Los metales son lixiviados de rocas sedimentarias y/o volcánicas. También se muestran los diferentes estilos de mineralización y la posible relación con depósitos más profundos de óxido de hierro-apatito (IOA por sus siglas en inglés) (extraído de Barra et al., 2017).

La mayoría de los yacimientos de IOCG presentan varias características comunes, una de ellas es la mineralización de Cu/Au está estrechamente asociada con la alteración K (biotita y/o feldespato K) (Chen et al., 2013). Del Real et al. (2018) en el Distrito Candelaria-Punta del Cobre identifica un evento de mineralización (~115-110 Ma) dominado por alteración de K-Fe a Ca-K-Fe caracterizada por magnetita-actinolita-biotita-feldespato $\pm$ cuarzo, con mineralización de calcopirita-pirita $\pm$ pirrotita.

Modelos Alternativos Basados en las Principales Fuentes de Fluidos

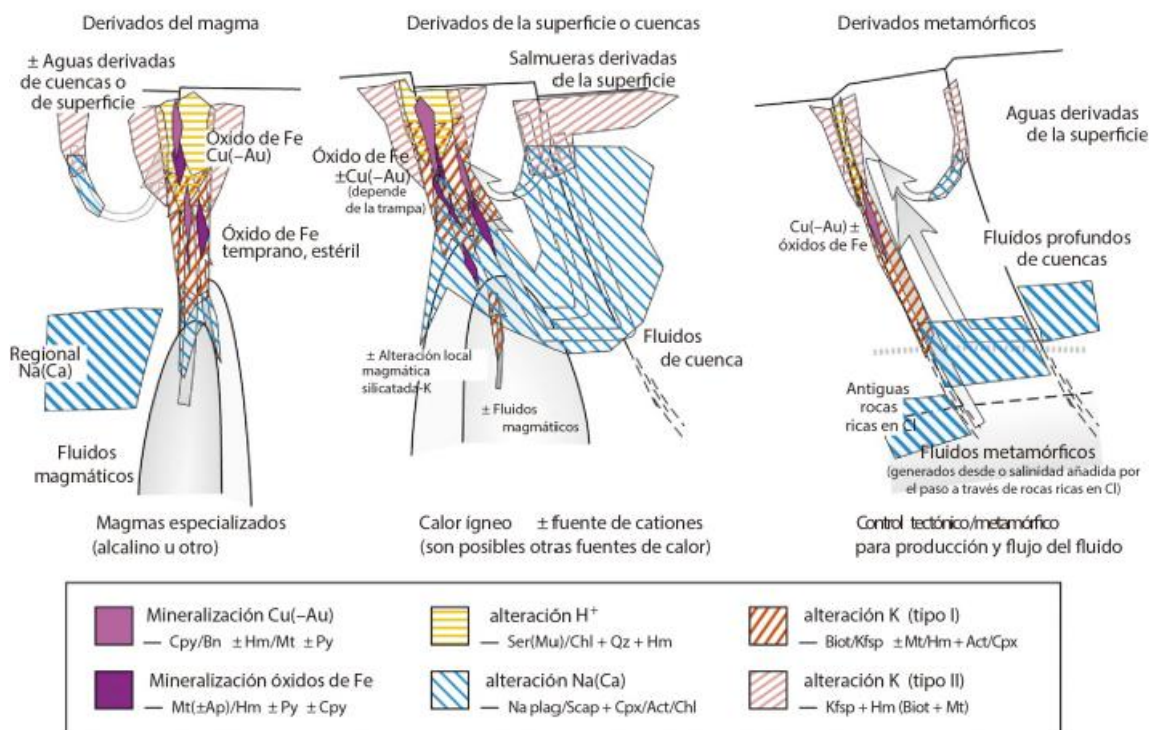


Figura 2.12. Arquitecturas y orígenes hidrotermales alternativos para sistemas IOCG ilustrando posibles fuentes de fluidos, recorridos del flujo hidrotermal, y distribución de alteraciones y menas (extraído de Morales, 2017, modificado de Barton, 2014).

En la proximidad de los cuerpos mineralizados las condiciones variables de presión y temperatura de la alteración y mineralización, se presentan en un espectro de depósitos que van desde aquellos en los que el óxido de Fe predominante es la magnetita y la alteración está representada por minerales como la biotita, el feldespato K y anfíbol; hasta los sistemas dominados por la hematita donde los principales minerales de alteración son la sericita y la clorita (Zandiyyeh et al., 2016; Richards et al., 2017; Burnham, 1979).

La zona distal y superficial de los sistemas IOCG está dominada por la hematita, mientras que la magnetita se forma antes y a mayor profundidad. Los metales que no son precipitados en estos entornos pobres en S y parcialmente oxidados podrían generar halos distales de Zn, Mn o Ag-Co-U (Zandiyyeh et al., 2016).

Generalmente se asocian estos depósitos a uno o pocos eventos magmáticos alcanzando volúmenes moderados (pocos Km²) de alteración Ca-Na habitualmente proximal y alteración ácida (hidrólisis) corrientemente distal y puede contener mineralización, pero en menor

cantidad (Oberreuter, 2017). Por lo que un área favorable debería estar sobre o próxima a una cámara magmática y tener una zona con metasomatismo intenso de Si, K y Fe, generalmente magnetita o hematita para sistemas de menor temperatura (Barton y Johnson, 2000).

2.8.3 Distribución temporal

Diversos autores han planteado que la formación de depósitos IOCG en el Precámbrico coincide con las etapas de ensamble y desmembramiento de los supercontinentes de Kenorland, Columbia y Rodinia (Kerrick et al., 2000; Goldfarb et al., 2001).

Los depósitos de óxido de hierro o de magnetita-apatita muestran un lapso similar (cerca de 300-800 Ma) a la de los depósitos de IOCG, pero se cree que se formaron predominantemente durante los periodos de orogénesis convergente (Groves et al., 2010). La Figura 2.13. muestra la distribución temporal de los depósitos de IOCG *sensu stricto*, junto con los yacimientos de óxido de hierro (P). Se puede observar que, en el Precámbrico, las provincias gigantes de IOCG se formaron entre 100 y 200 Ma después del ensamblaje del supercontinente (Groves et al 2005a, b, c; Franklin et al., 2005). Cabe señalar que no se conocen yacimientos de IOCG anteriores al primer supercontinente, Kenorland, y el mayor yacimiento conocido se encuentra en el primer supercontinente extenso, Columbia (Groves y Bierlein, 2007). Esto es coherente con un entorno intracratónico para la mineralización, muy probablemente relacionada con el hundimiento del manto (Kerrick et al., 2005, 2008).

2.8.4 Tectónica

Una característica para la interpretación del origen de estos depósitos es su tamaño, muchos de estos depósitos pueden contener más de 100 millones de toneladas de rocas mineralizadas (Hitzman et al., 1992). Lo que sugiere que tuvieron que estar implicados grandes volúmenes de fluidos hidrotermales a una temperatura elevada para su formación (Oreskes, 1990; Oreskes y Einaudi, 1992; Tappert et al., 2013). Las fracturas generadas por la extensión pueden proporcionar los canales por los cuales estos fluidos pueden llegar a niveles poco profundos de la corteza (Norton, 1982; Sleep, 1983; Groves et al., 2010).

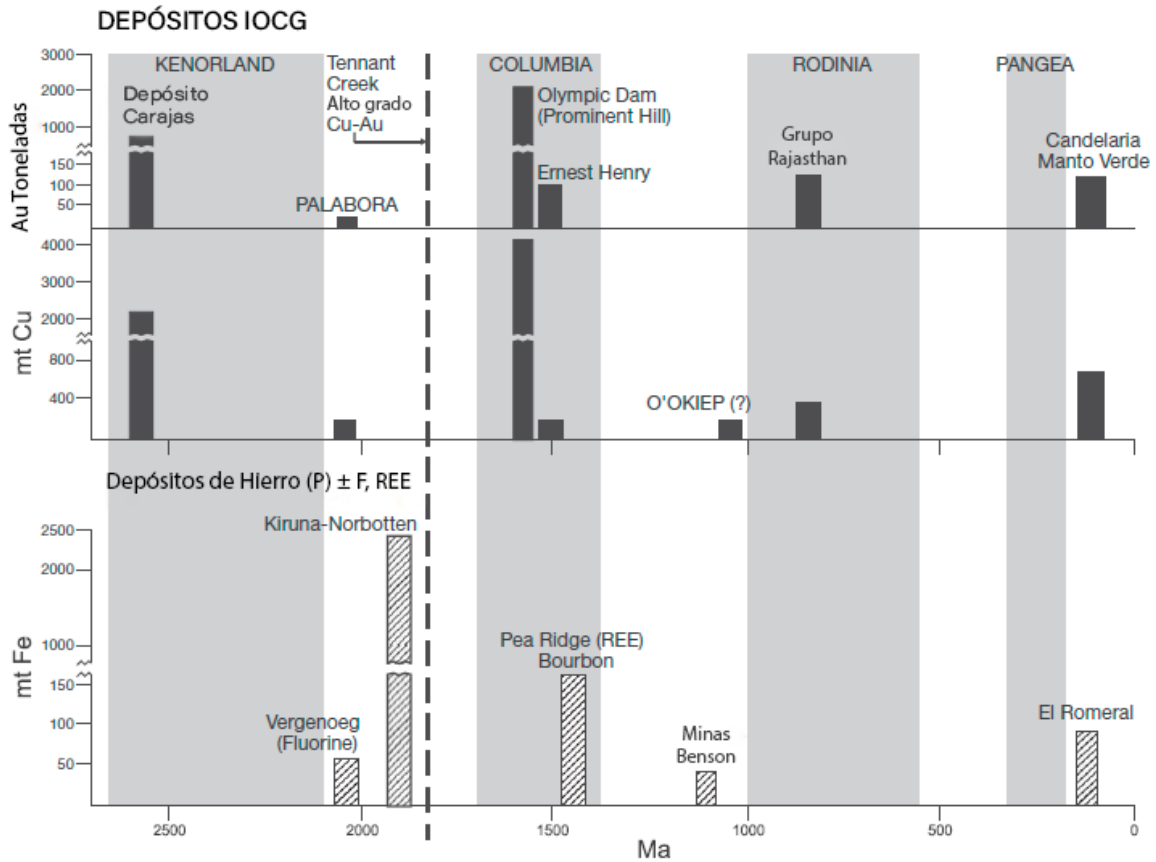


Figura 2.13. Distribución temporal de los grandes depósitos de IOCG *sensu stricto* y óxidos de hierro (P, F, REE) en función del ciclo de supercontinentes. Los periodos más comúnmente aceptados tras el ensamblaje de los supercontinentes y antes de la ruptura están sombreados. Las fechas de formación y ruptura de los supercontinentes son aproximadas (extraído de Groves et al., 2010).

Los depósitos de IOCG pueden formarse en diferentes ambientes tectónicos, como en entornos anorogénicos e intracratónicos relacionados con magmas derivados del manto de origen profundo del Arcaico o del Paleoproterozoico (Hand et al., 2007; Sun y McDonough, 1989), o en ambientes de trasarco reactivados relacionados con la subducción de márgenes continentales (Bindeman y Bailey, 1999; Moss et al., 2001; Haywood, 2008; Groves et al., 2005a, b). En estos casos, los fundidos parciales pueden ser responsables de la mineralización. El distrito Candelaria-Punta del Cobre es un ejemplo de mineralización relacionada con magmatismo de arco volcánico, extensión y formación de cuencas de trasarco (Del Real et al., 2018; Arévalo et al., 2006). Un modelo esquemático de la generación de depósitos de IOCG se muestra en la Figura 2.14.

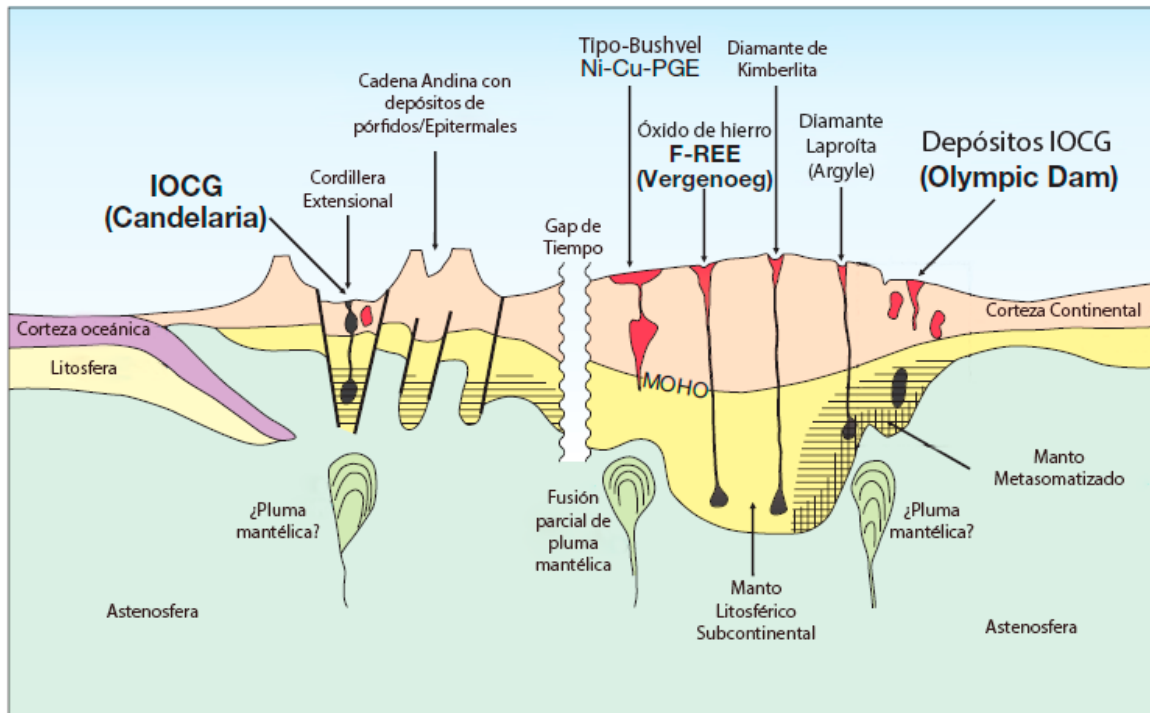


Figura 2.14. Diagrama esquemático que muestra el entorno tectónico y litosférico de los depósitos de IOCG, tanto en los cratones precámbricos como en las partes extensionales de los arcos cordilleranos. Obsérvese que los yacimientos de IOCG se forman en la corteza continental a bordo de los márgenes de los cratones sobre la litosfera del manto metasomatizado, normalmente de edad Arcaica. El espesor de la corteza está exagerada con respecto a la litosfera para acomodar los detalles (Extraído de Groves et al., 2010).

2.8.5 Relación genética depósitos IOA e IOCG

La clasificación y génesis de la mineralización en los IOCG se ha vuelto cada vez problemática, sobre todo desde que se han ido añadiendo diversos miembros. Por ejemplo, esta complejidad se ilustra en incluir dentro de los IOCG, sistemas de mineralización pobres en Cu-Au como los tipos Kiruna y El Laco, originalmente clasificados como depósitos magnetita-apatita de origen magmático (Geijer, 1931; Frietsch, 1978; Park, 1961). A medida que fueron identificados más yacimientos de magnetita-apatita de tipo Kiruna, y que estudios geoquímicos detallados permitieron mejorar la discriminación de las generaciones magmáticas e hidrotermales de los óxidos de Fe (Dare et al., 2014; Knipping et al., 2015a), estos depósitos se consideran afines a los sistemas IOCG, y se reclasificaron como un subgrupo distinto ampliamente desarrollados, el tipo apatita de óxido de hierro (IOA) (Geijer, 1931; Dmitrijeva et al., 2019b).

Diversos autores han propuesto la existencia de un vínculo genético entre los sistemas IOCG e IOA, ya que son muy similares en cuanto a la configuración tectónica, la mineralización y la alteración (Hitzman et al., 1992). Además, en muchos casos se ha sugerido que estos dos tipos de depósitos son parte del mismo sistema hidrotermal, donde los IOA corresponderían a la parte más profunda y estéril en Cu(-Au) de algunos sistemas IOCG (Espinoza et al., 1996; Naslund et al., 2002; Sillitoe, 2003; Barton, 2014, 2015; Knipping et al., 2015a; Bilenker et al., 2016).

Se ha propuesto un modelo magmático para la formación de depósitos de IOA en el que la magnetita se cristaliza a partir de fundidos de óxido ricos en volátiles (Nystroem y Henriquez, 1994; Velasco et al., 2016). Sin embargo, la asociación magnetita-apatita también forma parte de la alteración Ca-Na temprana dentro de los sistemas de IOCG y no es una discriminación entre los dos tipos de depósitos (Ismail et al., 2014; Krneta et al., 2016, 2017a, 2017b). La falta de alteración alcalina (Na-K) es una característica discriminatoria para algunos depósitos de IOA (Knipping et al., 2015b).

La presencia de sulfuros menores en los depósitos de IOA, y de magnetita-actinolita temprana anterior a la mineralización de Cu en los depósitos de IOCG, apuntan a una transición entre los estilos de IOA y de mineralización de IOCG poco profunda (Chen et al., 2010; Benavides et al., 2007; De Haller y Fontboté 2009; Lopez et al., 2014; Veloso et al., 2019). Sin embargo, estos autores asumen que tanto la mineralización de IOA como de IOCG eran el resultado de un único fluido de enfriamiento.

Rodriguez-Mustafa et al. (2020), mediante la caracterización geoquímica y textural de la magnetita y la actinolita del depósito Candelaria y el prospecto IOA Quince identificó un origen magmático/magmático-hidrotermal. Sugiere que la parte más profunda de Candelaria representa la transición entre la mineralización somera de IOCG y la IOA más profunda análogo al prospecto de IOA de Quince, apoyando la hipótesis de que los depósitos de IOA e IOCG son partes de un único continuo genético. Del Real et al. (2021) propone que analizando las variaciones químicas de la actinolita se podrá observar la naturaleza del fluido que precipita el mineral, entendiendo los cambios entre las etapas de pre-mineralización y mineralización, además, de señalar que estos depósitos pueden ocurrir sobreimpuestos.

CAPÍTULO 3: MARCO GEOLÓGICO

Las áreas de estudio, el Distrito Candelaria-Punta del Cobre y el Distrito Mantoverde, se ubican en la Cordillera de la Costa de la región de Atacama (Figura 3.1). En este capítulo se describen los antecedentes geológicos relevantes para este trabajo.

3.1. Marco geológico regional

3.1.1. Geología regional de la Cordillera de la Costa

En la Cordillera de la Costa de Atacama (CCA) el basamento está formado por rocas metasedimentarias del Devonico-Carbonífero, como la Formación Las Tórtolas, la cual está constituida por una secuencia turbidítica distal con intercalaciones menores de calizas, chert pelágicos, conglomerados y metabasitas; y el Complejo Epimetamórfico Chañaral que corresponde al equivalente deformado de la Fm. Las Tórtolas (Bell, 1982; 1984; Berg y Baumann, 1985; Lara y Godoy, 1998; Arévalo, 2005). La Formación La Negra corresponde a una gruesa sucesión volcánica-volcanoclástica, datada entre Jurásico Medio y Jurásico Tardío (Charrier et al., 2007; Lara y Godoy, 1998; García, 1967).

Entre los 26°S-30°S, al este de Chañaral se registra la Formación Sierra Fraga de edad Jurásico Medio-Tardío, la cual es una extensión distal de la Fm. La Negra, y consiste en una sucesión de lavas andesíticas, basalto-andesitas y rocas sedimentarias detríticas y calcáreas (Sepúlveda y Naranjo, 1982; Iriarte et al., 1996). Durante el Jurásico Tardío-Valanginiano el arco de La Negra es reemplazado gradualmente por la Formación Punta del Cobre denotando una actividad volcánica continua durante el Jurásico al Cretácico Temprano. La Fm. Punta del Cobre consiste en una sucesión volcánica en su parte inferior (Miembro Geraldo-Negro) y volcanoclástico en su parte superior (Miembro Algarrobos) (Seagerstrom y Ruiz, 1962; Marschik y Fontboté, 1996; Ortiz et al., 1966), subyace a sedimentos carbonatados marinos someros del Grupo Chañarcillo al sur de Copiapó (Marschik y Fontboté, 2001b; Corvalán, 1974; Cisternas, 1986; Cisternas y Díaz, 1990; Pérez et al., 1990).

Al sur de Copiapó sobre la Fm. Punta del Cobre se encuentra el Grupo Chañarcillo el cual consiste en sedimentos carbonatados marinos someros y contenidos subordinados de rocas volcanoclásticas en la base y en el techo, de edad Valangiano Superior-Aptiano (Marschik y

Fontboté, 2001b; Corvalán, 1974; Cisternas, 1986; Cisternas y Díaz, 1990; Pérez et al., 1990); y sucesiones volcano-sedimentarias con intercalaciones sedimentarias detríticas y calcáreas marinas del Hauteriviano-Aptiano Temprano de la Formación Bandurrias, la cual corresponde a una zona de transición entre el arco al oeste y la cuenca de trasarco al este (Figura 3.1) (Segerstrom, 1960; Jurgan, 1977a, b). La Formación Cerrillos al norte de la Quebrada Paipote sobreyace en concordancia al Grupo Chañarcillo, y también a subyacente Fm. Bandurrias. La Fm. Cerrillos está compuesta por una secuencia volcanosedimentaria predominantemente continental de edad Albiano-Santoniano (Segerstrom y Parker, 1959).

El magmatismo del Jurásico-Cretácico temprano produjo abundantes cuerpos intrusivos en la CCA incluyendo a los monzogranitos, granodioritas, tonalitas, y gabros del Plutón Flamenco (200-194 Ma), que intruye al Complejo Epimetamórfico Chañaral (Grocott et al., 1994). El Plutón Las Ánimas (160-150 Ma), de composición diorítica, microdiorítica y tonalítica, Al este de la CCA se encuentran plutones del Cretácico Inferior, e.g. el Plutón Cerro Moradito (155-140 Ma) de composición monzonitas cuarcíferas, el Plutón Cerro Morado (140-137 Ma) de composición tonalítica a diorítica, y el Plutón la Borracha (114-103 Ma) constituido de monzodioritas cuarcíferas, monzonitas y dioritas (Arévalo, 2005; Dallmeyer et al., 1996; Lara y Godoy, 1998).

Las rocas en el área de estudio se relacionarían a una zona de subducción de la placa de Aluk bajo el continente sudamericano, la que provocó el desarrollo de un arco magmático con orientación N-S paralelo al margen continental y la formación de una cuenca de trasarco marina asociada al magmatismo del Jurásico – Cretácico Temprano, recibiendo aportes volcánicos y facies sedimentarias principalmente carbonatadas (Mpodozis y Ramos 1990; Scheuber y Andriessen 1990; Coira et al., 1982; Dalziel et al., 1987; Stern et al., 1991; Atherton y Aguirre, 1992; Oliveros et al., 2020).

Durante el Cretácico Temprano, el sistema arco-cuenca sufrió importantes intrusiones de dimensiones batolíticas, con cuerpos intrusivos de composiciones principalmente dioríticas y granodioríticas (Peréz, 2017). También durante este periodo, ocurre la migración del arco hacia el este; inversión tectónica de las cuencas de trasarco y el desarrollo de cuencas de antepaís (Charrier et al., 2007).

Los depósitos en esta franja metalogénica se encuentran fuertemente controlados por el Sistema de Falla de Atacama (Loyola, 2016; Scheuber y González, 1999). La actividad plutónica se desarrolla bajo condiciones transtensionales y en contemporáneo con fallas de rumbo sinistral a lo largo del Sistema de Falla de Atacama (Arabasz, 1971; Richards et al., 2017). Contemporáneo a este evento estaría el origen del Cinturón de Hierro Chileno, el cual es una zona de gran importancia metalogénica, extendiéndose desde el sur de Perú hasta la Región de Coquimbo de Chile (Loyola, 2016).

3.1.2. Geología Estructural

3.1.2.1. Sistema de Falla Atacama

El Sistema de Falla Atacama (SFA) se extiende a lo largo de la Cordillera de la Costa por más de 1000 km desde Iquique hasta La Serena (Naranjo, 1987; Thiele y Pincheira, 1987; Brown et al., 1993). La geometría del SFA corresponde a un sistema de fallas de grandes dimensiones interconectadas, paralelas al arco con orientación predominantemente N-S, y otras fallas subsidiarias de orientación NW formando dúplex a escala regional y local (Charrier et al., 2007). A su vez registra una compleja evolución cinemática que comienza en el Jurásico Temprano y que predomina en el Cretácico Temprano con desplazamientos dúctiles sinistral transtensionales pasando a una deformación frágil sinistral transpresional en el Aptiano Tardío-Albiense (Brown et al., 1993; Scheuber y Gonzalez, 1999; Cembrano et al., 2005). Se ha documentado actividad del SFA durante el Cenozoico en el cual el componente fue principalmente vertical. Esta actividad dio lugar a rasgos morfológicos, especialmente bien preservados en las latitudes de Taltal, Antofagasta y Mejillones. Dentro de los rasgos morfológicos se destaca el levantamiento del lado occidental de la falla, el cierre de valles fluviales, presencia de terraza y escarpes. En el Mioceno el SFA acomodó el levantamiento relativo del lado occidental de la CCA en la zona de Chañaral.

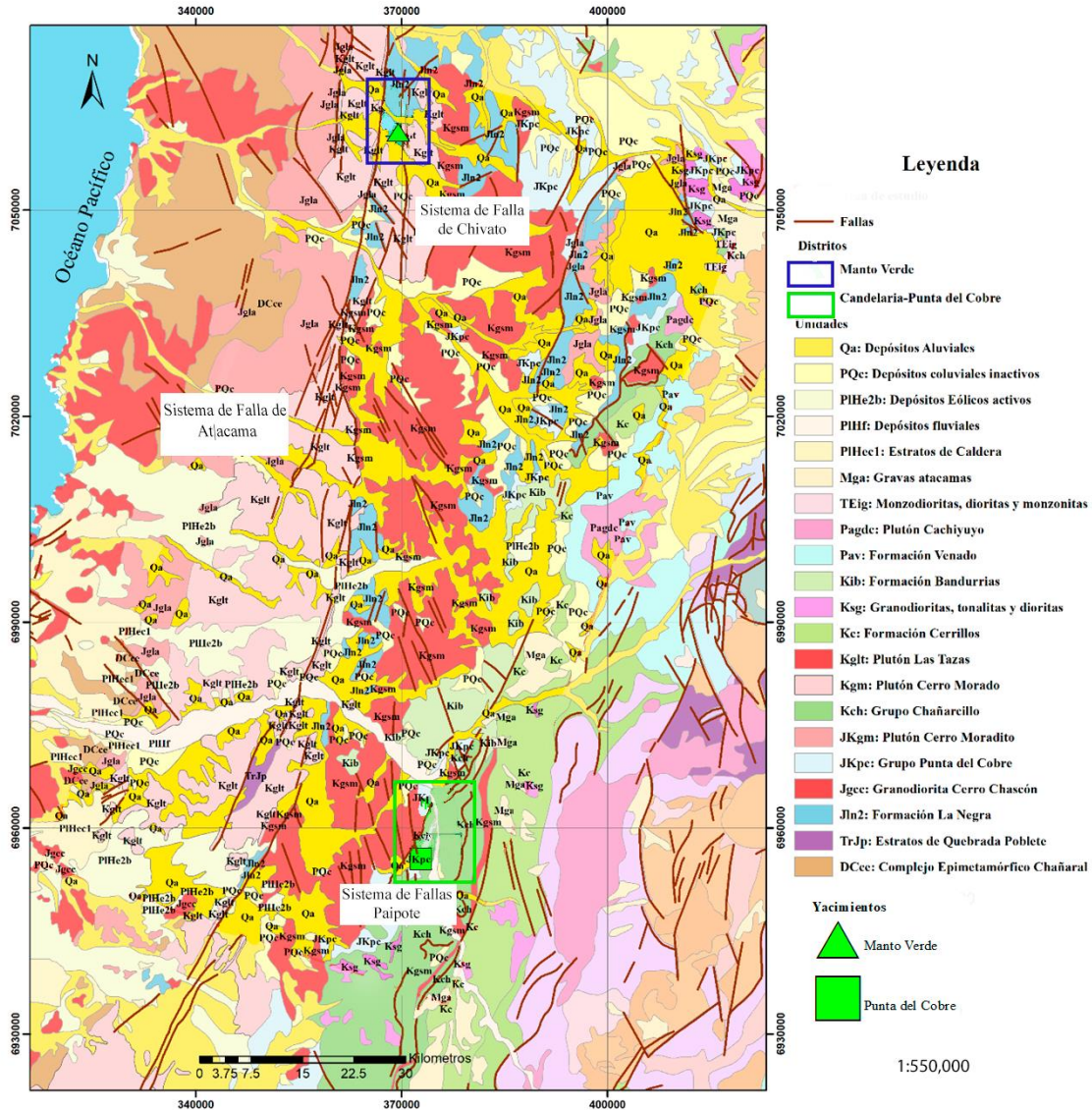


Figura 3.1. Mapa geológico para el segmento de la Cordillera de la Costa donde se encuentran los distritos Candalaria-Punta del Cobre y Mantoverde. Geología de SERNAGEOMIN (2003).

El SFA jugó un rol importante en el control del emplazamiento de plutones del Jurásico Tardío y Cretácico Temprano (Grocott et al., 1994; Grocott y Taylor, 2002). Además, es el control estructural primario en el emplazamiento de los depósitos tipo IOCG como Mantoverde y los depósitos de magnetita pobres en sulfuro del Cinturón de Hierro Chileno (Figura 3.2) (Sillitoe, 2003; Creixell et al., 2009; Richards et al., 2017; Espinoza, 1990; Grocott y Taylor, 2002; Herrera et al., 2007; Brown et al., 1993; Scheuber y González, 1999; Arévalo et al., 2003).

3.1.2.2. Sistema de Falla Chivato

El Sistema de Falla Chivato (SFC) se ubica en el lado este de la CCA y corresponde a un sistema extensional de fallas oblicuas con componente dextral y rumbo NE-NNE sin-plutónicas del Triásico Tardío-Cretácico, la cual registra una sucesión de episodios magmáticos y de mineralización (Arévalo et al., 2006; Godoy et al., 1997). En el SFC las zonas de cizalla dúctil de alta temperatura están ligadas con el emplazamiento de plutones de arco, y se asemejan al SFA en las condiciones de mineralización tipo magnetita-apatito durante la transición dúctil-frágil en estas zonas de fallas (Bonson, 1998; Grocott y Taylor, 2002).

Godoy et al. (1997) segmenta al SFC en dos tramos uno llamado Segmento Chivato (septentrional) con fuerte manteo hacia el este, y otro segmento llamado Agua del Sol con transcurrancia sinistral con manteo hacia el oeste. Aproximadamente durante el Cretácico Tardío, el SFC se invirtió significativamente por la transpresión sinistral de un sistema de transferencia NW reactivado en el límite de la Cordillera de la Costa-Precordillera (Godoy et al., 1997; Arévalo, 1999; Grocott y Taylor, 2002). El acortamiento ocurrió durante el emplazamiento de un conjunto de intrusiones graníticas del Cretácico Tardío en la Precordillera hace aproximadamente 69-65 Ma (Gipson, 2002).

3.1.2.3. Sistema de Falla Paipote

El Sistema de Falla Paipote (SFP) está ubicado en el contacto del batolito de Copiapó y las rocas de la Formación Punta del Cobre-Grupo Chañarcillo, consiste en una serie de fallas de empuje con vergencia oeste con sentidos de desplazamientos tanto normales como inversos que se formaron durante la inversión de mediados del Cretácico (Arévalo et al., 2006; Del Real et al., 2018). Arévalo et al (2006) interpreta a la falla maestra Paipote como una falla invertida de límite de cuenca debido al cambio de facies (Grupo Chañarcillo-Formación Bandurrias) y de espesor a lo largo del Cabalgamiento de Paipote.

Hacia el sur, el SFP forma el límite entre los complejos plutónicos y las secuencias estratificadas. Esto implica que la falla se reactivó a mediados del Cretácico durante el emplazamiento de intrusiones plutónicas (Arévalo et al., 2006; Grocott y Taylor, 2002). El SFP corta a las unidades Fm. Punta del Cobre, Fm. Bandurrias, con excepción de las formaciones superiores del Grupo Chañarcillo y la Fm. Cerrillos (Del Real et al., 2018).

La Faja Plegada y Corrida de Paipote es un sistema estructural de tipo flor positiva, compuesto de fallas frágiles de alto ángulo y movimiento inverso con componente sinistral, enraizada en el SFP, que desplaza y deforma las secuencias del Cretácico Temprano a lo largo de la Quebrada Paipote y del valle del río Copiapó, y genera un anticlinal, conocido como Anticlinorio de Tierra Amarilla (Lara y Godoy, 1998; Arévalo et al., 2006; Martínez et al., 2013). La principal estructura del sistema es el Cabalgamiento de Paipote que tiene una inclinación de 20° a 80° al noroeste. El anticlinal de Tierra Amarilla está presente en la mina Candelaria, involucrando tanto al yacimiento como a la roca hospedante (Arévalo et al., 2006; Grocott y Taylor, 2002).

3.1.3. Cinturón de Hierro Chileno

El Cinturón de Hierro Chileno (CHC) o Franja Ferrifera se encuentra entre los 25°30' y 31°S, y está formado por más de 40 depósitos de hierro de importancia económica. El eje de la franja está gruesamente definido por el SFA siguiendo una tendencia general NNE en un área de 600 km de largo y 30-40 km de ancho, paralela al eje de la fosa peruano-chilena (Ménard, 1995).

Los depósitos del CHC están mayoritariamente hospedados en conglomerados, brechas o flujos volcanoclásticos andesíticos del Jurásico Tardío, y en rocas intrusivas cretácicas del batolito costero chileno (Berg y Baumann 1985; Berg y Breitzkreuz 1983; Dallmeyer et al. 1996; Gelcich et al. 2005; Marschik y Söllner 2006; Wilson et al. 2000; Lara y Godoy 1998). En el CHC se encuentran varios yacimientos tipo magnetita-apatito (Los Colorados y Carmen), IOCG (Mantoverde), y algunos pórfidos de Cu-Au (Andacollo) (Loyola, 2016; Aguilef, 2012). Los depósitos de IOA presentes en el CHC se caracterizan por presentar una mineralización dominada por la asociación magnetita-apatito, con mayor predominancia de magnetita; y actinolita como principal mineral de alteración (Los Colorados, Carmen). Además, pueden contener cantidades menores o casi ausencia de pirita, calcopirita, bornita, y en algunos casos pirrotina, covelina y digenita (Cerro Imán). Estos depósitos presentan alteración hidrotermal predominantemente propilítica, argílica, Ca-Na y actinolización (Fredes, 2017; Munizaga y Lagos., 2015; Velasco y Tornos, 2009; Vivallo et al., 1995; Treloar y Colley, 1996).

Los yacimientos IOCG como Mantoverde y/o Candelaria y varios depósitos IOA del CHC están hospedados por fallas pertenecientes al SFA (Espinoza, 1990; Grocott y Taylor, 2002). En otras áreas los yacimientos IOCG están relacionados espacial y temporalmente por zonas de falla de rumbo sinistral NW y WNW (e.g. Candelaria, Punta del Cobre, Carola; Arévalo et al., 2006; Del Real et al., 2018).

3.1.4. Depósitos IOCG en la región de Atacama

Entre los 26° y 29°S los depósitos de IOCG ocupan una distancia longitudinal de casi 700 km en el CHC. La mayoría de estos depósitos están alojados en la Fm. La Negra y a sus equivalentes estratigráficos más al sur (Ruiz et al., 1968; Bookstrom, 1977; Sillitoe, 2003).

Los principales depósitos de IOCG del CHC comprenden variadas combinaciones de mineralización tales como brechas, zonas de stockwork, skarns calcáreos y mantos de reemplazo, además de vetas, siendo estas últimas las más abundantes, tal como Candelaria-Punta del Cobre, Mantoverde y Carrizal Alto (Martin et al., 1997; Marschik y Fontbote, 2001b; Vila et al. 1996; Zamora y Castillo 2001; Vidal et al. 1990; de Haller et al. 2002). Los depósitos de IOCG del CHC suelen ser subdivididos en dos subtipos, los dominados por magnetita y los dominados por hematita, ambos subtipos presentan vetas con calcopirita ± bornita (Ruiz y Ericksen, 1962; Ruiz et al., 1965).

Tanto las vetas de IOCG ricas en magnetita como en hematita especular contienen calcopirita y generalmente pirita subordinada, pero en unos pocos casos la bornita acompaña a la calcopirita. Ambos subtipos de IOCG pueden contener cobaltita, saflorita, danaita (todas con Co y As), niccolita, cloantina (ambas con Ni y As), molibdenita y uranita (Ruiz et al., 1965).

La alteración relacionada con los depósitos de IOCG es compleja y variada. La alteración Na o Ca-Na, caracterizada por albita con o sin actinolita, se da en algunos de los distritos de IOCG (es el caso de, Candelaria-Punta del Cobre; Marschik y Fontboté, 1996; 2001b), pero aparentemente está ausente en otros lugares (por ejemplo, Mantoverde; Vila et al., 1996; Cornejo et al., 2000). Las vetas ricas en magnetita contienen estrechos halos de alteración que contienen actinolita, biotita, albita, feldespato K, epidota, cuarzo clorita, sericita y escapolita (Ruiz et al. 1965; Boric et al. 1990; Espinoza et al. 1996). En cambio, las vetas ricas en hematita tienden a contener halos de alteración caracterizados por sericita y/o clorita.

Los depósitos IOCG de los Andes Centrales han sido ampliamente estudiados, y se han propuestos distintos modelos para la formación de estos depósitos basados en la interpretación de datos de varios sistemas diferentes. Los modelos han conjurado diferentes fuentes de fluido (Del Real, 2020):

- (1) Salmueras salinas oxidadas derivadas de evaporitas que se formaron en cuencas de trasarco continental durante el Jurásico Superior y el Cretácico Inferior, lo que tendría relación con las evidencias paleoclimáticas de condiciones áridas a semiáridas en la región desde el Mesozoico (Suárez y Bell 1987). Además, la presencia de evaporitas en secuencias del Triásico Tardío al Cretácico Temprano (Suárez y Bell 1987; Hopper et al., 2000), las cuales podrían aportar sulfatos y/o cloro a los fluidos de la corteza terrestre (Mathur et al. 2002).
- (2) Fluidos magmático-hidrotérmicos que evolucionaron a partir de intrusiones ígneas asociadas temporal y espacialmente, correlacionándose con la conexión genética entre los depósitos IOCG de los Andes centrales y magmas gabrodioríticos a dioríticos propuesto por Sillitoe (2003).

Los mayores depósitos IOCG pudieron ser originados por el ingreso de fluidos de mena en secuencias volcano-sedimentarias relativamente permeables. En consecuencia, la mayoría de estos depósitos se encuentran asociados estratigráficamente a formaciones del Jurásico (Formación La Negra) y Cretácico (Formación Punta del Cobre) (Sillitoe, 2003; Del Real et al., 2018).

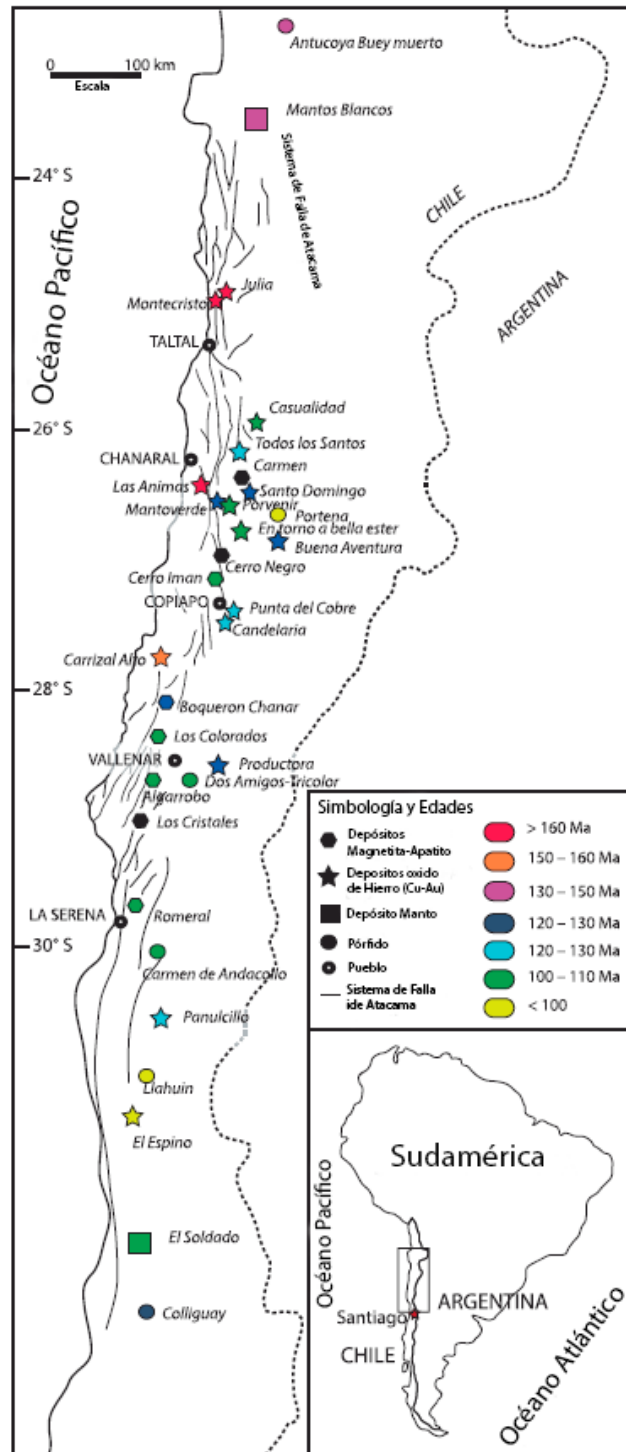


Figura 3.2. Mapa de localización de los depósitos de IOCG, IOA y pórfidos formados durante el Jurásico Tardío – Cretácico Temprano en el norte de Chile (extraído de Del Real et al., 2018).

3.1.5. Evolución y relación genética de los depósitos IOCG e IOA de la región de Atacama

Como se discutió en el capítulo anterior el vínculo genético IOA e IOCG no es del todo claro. Si los depósitos de IOCG e IOA representaran diferentes partes de la evolución de un único sistema hidrotermal, entonces cualquier modelo genético viable que explique un tipo de depósito debería explicar también el otro tipo (Childress et al., 2020). Diversos modelos propuestos tienen en común varias características, tales como la mineralización de la magnetita y sulfuros, alteración calco-sódica y una relación espaciotemporal con intrusivos de composición diorítico a granodiorítica (Knipping et al., 2015b; Reich et al., 2016; Barra et al., 2017, Rojas et al., 2018), existe una convergencia sobre la zonificación mineral vertical y la influencia de la combinación y/o enfriamiento de un fluido magmático inicial a un fluido magmático-hidrotermal (Espinoza et al., 1996; Palma et al., 2021; Childress et al., 2020; Barra et al., 2017; Huang et al., 2019).

En términos generales, estos modelos sugieren que el fluido magmático inicialmente daría origen, a mayor profundidad, a los depósitos IOA (con mineralización de magnetita-apatita-actinolita, pero con bajos contenidos en Cu-Au), y que posteriormente puede seguir ascendiendo y enfriándose hasta convertirse y/o mezclarse con un fluido magmático-hidrotermal hacia la parte más somera y formar depósitos IOCG (rico en magnetita-hematita, sulfuros de cobre y presencia de carbonatos como la calcita) (Espinoza et al., 1996; Knipping et al. 2015a; Palma et al., 2021; Barra et al., 2017; Barton 2013).

La hipótesis de una conexión genética entre los depósitos IOCG y IOA ha sido respaldada por estudios realizados en depósitos IOA (Los Colorados, Carmen y Fresia; ver: Huang et al., 2019; Palma et al., 2021; Treloar y Colley, 1996; Knipping et al., 2015a) y depósitos IOCG (La Estrella, Candelaria, Mantoverde y Diego de Almagro; ver: Childress et al., 2020; Neumann et al., 2017; Benavidez et al., 2007). Se ha propuesto que el régimen tectónico extensional jugó un rol importante en la génesis de los depósitos IOA en la región de Atacama, donde las fallas principales habrían permitido el ascenso de fluidos magmáticos (p.ej. Knipping et al., 2015b). Por su parte, los depósitos IOCG estarían asociados a fallas secundarias subordinadas a las fallas principales, en las cuales los fluidos magmático-hidrotermales e hidrotermales habrían podido precipitar magnetita y hematita, sulfuros de Cu-(Fe) y Au (Arévalo et al., 2006; Benavides et al., 2007; Knipping et al., 2015a, b;

Childress et al., 2020; Palma et al., 2021; Barra et al., 2017; Varela, 2021). La interpretación propuesta para la evolución de los fluidos mineralizadores en ambientes IOA-IOCG es consistente con la dependencia termal de la solubilidad del azufre y metales. Datos de transporte de metales por vapor de los sistemas de formación de depósitos magmáticos-hidrotermales (Williams-Jones y Heinrich, 2005) y estudios experimentales de la solubilidad de los metales, los cuales consistieron en la recopilación de datos de distintas fuentes geológicas como fumarolas relacionadas a depósitos minerales, gases volcánicos y sistemas geotermiales obtenidos en Nicaragua, Indonesia, Rusia y México; como también de datos conseguidos mediante experimentos (Chou y Eugster 1977; Wood y Samson 1998; Simon et al. 2004; Hurtig y Williams-Jones 2014; Williams-Jones y Migdisov 2014), demuestran que el fluido magmático-hidrotermal se enriquece en S y en metales como el Fe, Cu y Au a medida que se enfría a temperaturas menores a 500°C (Childress et al., 2020).

A lo anteriormente señalado se unen otros estudios que evidencian la influencia de fluido magmático a magmático-hidrotermal y zonificación mineral de los depósitos IOA e IOCG en la región de Atacama. Los datos químicos de los elementos traza de la magnetita y hematita, junto con las temperaturas de homogenización de las inclusiones fluidas en Mantoverde publicadas por Rieger et al. (2012), y los datos texturales y composicionales de la magnetita en yacimientos IOCG e IOA en Brasil, Chile, Australia, Suecia, Canadá y USA obtenidos por Huang et al. (2019), muestran un núcleo de magnetita de origen ígneo y un borde de origen magmático-hidrotermal. Además, los datos de enriquecimiento de elementos traza en los núcleos y en una fase posterior en los bordes con inclusiones fluidas de halita en magnetita de los depósitos de Mantoverde y Los Colorados, y los datos de termometría de la magnetita en los depósitos andinos IOA de Palma et al. (2021), son consistentes con la zonificación propuesta por Espinoza et al. (1996) y Knipping et al. (2015a y 2015b).

3.1.6. REE en Atacama

Estudios han informado de varios cientos de ppm de LREE en el distrito Candelaria-Punta del Cobre, al menos en parte en allanita (Marschik et al. 2000, 2003; C. Osterman en Ray et al., 2002; Del Real et al., 2018). Los valores normalizados con respecto a la condrita REE para las diferentes fases intrusivas del batolito de Copiapó muestran un ligero enriquecimiento en LREE, ningún agotamiento en MREE, y una ligera o nula anomalía

negativa en Eu (Del Real et al., 2018; Marschik et al., 2003). El conjunto de diques de dacita datados por Del Real et al. (2018) tienen firmas LREE y MREE similares a las del batolito, estos diques tienen una pronunciada anomalía negativa de Eu. Las rocas monzodioríticas y monzoníticas cuarzosas del distrito Candelaria - Punta del Cobre tienden a tener contenidos de LREE ligeramente elevados, y parecen tener una inflexión entre Sm y Eu en comparación con los patrones de las rocas dioríticas y tonalíticas (Marschik et al., 2000, 2003).

Rieger et al. (2010) analizaron los REE del distrito Mantoverde para localizar áreas o unidades enriquecidas en REE y evaluar posibles fuentes del REE. Aunque el distrito Mantoverde no muestra enriquecimiento extremo observado en yacimientos como Olympic Dam, se detectan algunas concentraciones significativas de REE (Oreskes y Einaudi, 1990). Las rocas huésped del distrito Mantoverde muestran una variedad de patrones de REE, incluyendo un carácter relativamente plano, de ligera a moderadamente enriquecido, enriquecidos en HREE respecto a LREE o viceversa; patrones con una anomalía de Eu positiva o negativa más o menos pronunciada; o sin anomalía de Eu. Parte de los LREE están alojados en allanita y apatita, los cuales solo están presentes en pequeñas cantidades. Las concentraciones elevadas de LREE se dan localmente en el distrito Mantoverde, pero son atribuidas a la redistribución de los LREE dentro de los depósitos más que a la influencia de fuentes externas (Rieger et al., 2010a). No hay evidencia de una introducción significativa de LREE desde fuera del yacimiento por fluidos mineralizadores, por lo que es probable que estos elementos se redistribuyeron dentro del yacimiento (Rieger et al 2009).

Los depósitos tipo IOA de la región de Atacama presentan una mayor concentración de REE a diferencia de los depósitos tipo IOCG, aunque sus valores no son económicamente rentables, y los métodos actuales presentan desafíos para su explotación (Rojas et al., 2018). En el depósito IOA Carmen del Hierro se desarrolló apatito y monacita (Treloar y Colley, 1996). La monacita se desarrolló como resultado de la exsolución del exceso del contenido en REE del flúor-apatito (Velasco y Tornos, 2009). Vivallo et al. (1995) mediante el método de activación neutrónica analizó muestras de rocas de los distritos Cerro Negro Norte, Cerro Imán y El Romeral, estos depósitos presentaron patrones relativamente horizontales de REE, con anomalías de Eu poco pronunciadas en Cerro Negro Norte y mejor desarrolladas en el distrito El Romeral y en algunas de las muestras de Cerro Imán.

Venegas (2015) mediante XRF portátil estudió el valor económico de REE en el depósito feldespático/pegmatítico Mina Maciso, ubicada en la CCA, al noroeste de Sierra El Roble. El autor encontró valores anómalos de REE, valores promedio y máximo por encima de las estimaciones de la abundancia de los REE en la corteza de acuerdo con McGill (1997), los minerales portadores corresponden a allanita-monazita, con firma geoquímica de LREE. Por ejemplo, el valor promedio de Ce es ~145 ppm y un máximo de 4000 ppm aprox. Muy por encima del rango de 20-46 ppm indicado por McGill (1997).

3.2. Geología Local

3.2.1. Distrito Candelaria–Punta del Cobre

El Distrito Candelaria-Punta del Cobre se encuentra ubicado en el borde oriental de la Cordillera de la Costa (27°30'S) (Figura 3.1) (Espinoza, 1985, 1990; Ménard, 1986; Ruiz y Peebles, 1988; Arévalo et al., 2006; Veloso et al., 2017). El Distrito se encuentra espacialmente asociado a batolitos dioríticos-granodioríticos del arco magmático del Cretácico Inferior (Sillitoe, 2003). La mineralización está hospedada en rocas volcánicas y volcanoclasticas del Jurásico-Cretácico Inferior, como la Formación La Negra, Formación Punta del Cobre y el Grupo Chañarcillo.

Estructuralmente, la zona corresponde a un cinturón altamente foliado con esquistos de biotita (Arévalo et al., 2006). La zona de Tierra Amarilla está marcada por un sistema de fallas sinplutónicas que registró una sucesión de episodios estructurales, magmáticos y de mineralización sinplutónicas del Cretácico Temprano (Arévalo, 1999). La Falla Paipote es uno de los rasgos estructurales principales del distrito, el cual se encuentra en el contacto entre el Batolito Copiapó con la Formación Punta del Cobre y el Grupo Chañarcillo. Esta falla se describe como una falla inversa formada durante la inversión del Cretácico (Arévalo et al., 2006; Veloso et al., 2017). En el distrito de Candelaria-Punta del Cobre, las fallas extensionales, iniciadas durante la formación de la cuenca sedimentaria que recubre los cuerpos mineralizados, se invirtieron al mismo tiempo que la actividad hidrotermal (Del Real et al., 2018).

3.2.1.1. Mineralización y alteración

La secuencia paragenética del distrito Candelaria-Punta del Cobre ha sido tratada por varios autores (Ryan et al., 1995; Martin et al., 1997; Arévalo, 1999; Ullrich et al., 1999; Marschik

y Fontboté, 2001a; Del Real et al., 2018, 2020). La mayoría de la mineralización del distrito Candelaria-Punta del Cobre está hospedada en la Formación Punta del Cobre, la cual informalmente se divide en cuatro partes (ver Figura 3.3): la más profunda corresponde a la unidad Andesita Inferior; encima en un mismo nivel estratigráfico en algunas zonas se observan la unidad Dacita y en otras zonas se observa un estrato volcano-sedimentario; por último, encima del nivel estratigráfico previamente mencionado se encuentra la unidad Andesita Superior (Del Real et al., 2018; Veloso et al., 2017; Arévalo et al., 2006).

La mineralización del distrito está alojada principalmente en andesitas masivas y brechas de la sección superior de la unidad Andesita Inferior y en las rocas basales de la unidad volcán-sedimentario suprayacente a la Andesita Inferior (Arévalo et al., 2006; Del Real et al., 2018, 2020). La mineralización consiste en (1) vetas de calcopirita-pirita-magnetita ($\pm$ pirrotita), (2) calcopirita-pirita-magnetita formando la matriz de rocas metavolcanicas biotitizadas y (3) calcopirita ($\pm$ pirrotita) en esquistos de biotita (Figura 3.4) (Arévalo et al., 2006; Marschik et al., 1997; Del Real et al., 2018, 2020; Marschik y Fontboté. 2001a). Las vetas de calcita y la carbonatización generalizada son comunes a niveles poco profundos en el distrito Candelaria-Punta del Cobre (Marschik et al., 2014). También se encuentran elevadas concentraciones de LREE, y trazas de molibdenita y arsenopirita (Mathur et al., 2002). El control de la mineralización en el distrito depende de varios factores como la paleopropundidad, fallas, litología y permeabilidad. Donde la permeabilidad de la roca huésped juega un papel principal, siendo la unidad volcano-sedimentario la unidad de mayor permeabilidad y presenta mineralización tipo “manto” (secciones estratigráficas horizontales ricas en magnetita) (Del Real et al., 2018, 2020).

El modelo paragenético más integrado es el propuesto por Marschik y Fontboté (2001a), el cual propone tres etapas principales de mineralización:

1. Etapa de óxido de hierro con un metasomatismo Na temprano y una etapa principal de metasomatismo de hierro y alteración potásica.
2. Etapa temprana de sulfuros representada por calcopirita-pirita con pirrotita ocasional como mineralogía principal del yacimiento
3. Etapa tardía de hematita y calcita.

Otros autores (e.g. Del Real et al., 2018) han agrupado las asociaciones en cinco eventos en lugar de tres, aunque la secuencia de eventos es relativamente similar al propuesto por Marschik y Fontboté (2001a). Las asociaciones minerales relacionadas con la alteración y la mineralización en los depósitos del distrito de Candelaria-Punta del Cobre reflejan la evolución de los fluidos y las reacciones fluido-roca que dan lugar a conjuntos minerales que han sido reconocidos en muchos otros distritos de IOCG (Hitzman et al., 1992; Williams et al., 2005; Del Real et al., 2018; Marschik y Fontboté. 2001a; Huang et al., 2019; Chen et al., 2010). La alteración en el distrito está mejor desarrollada hacia el oeste, y puede estar fuertemente vinculada a al emplazamiento del Batolito Copiapó (Ryan et al., 1994, 1995; Martin et al., 1997; Ullrich et al., 1999; Mathur et al., 2002; Marschik et al., 2000; Marschik y Fontboté 2001a). Los “mantos” presentes en el distrito se caracterizan por alteración pervasiva (Del Real et al., 2018, 2020).

La mineralización en el segmento más profundo del yacimiento Candelaria se concentra en una zona denominada “Brecha de magnetita”, la cual se caracteriza por presentar una intensa alteración de magnetita-actinolita-biotita-k-feldespatos y calcopirita-pirita diseminada y controlada por fracturas, posteriormente sobreimpresa por vetas y diseminaciones de magnetita-biotita-calcopirita-pirrotita $\pm$ pirita $\pm$ esfalerita $\pm$ mushketovita (Marschik y Fontboté, 2001a; del Real et al., 2018; Barton et al., 2013; 2015).

Dos grupos de temperatura registradas en actinolita por Del Real et al. (2021) indican que el yacimiento de Candelaria se formó durante dos etapas distintas de mineralización como parte de un sistema hidrotermal en evolución. Una primera etapa de alta temperatura (800-675°) rica en Fe y generalizada por abundante alteración magnetita-actinolita, y una segunda etapa de baja temperatura (550-700°C) más enriquecida en Cu. La primera etapa se correlaciona con la alteración temprana de magnetita-actinolita registrada en el distrito Candelaria-Punta del Cobre (Del Real et al., 2018). Las temperaturas de la segunda etapa coinciden con las temperaturas en las que el Cu alcanza su máxima solubilidad en un fluido hidrotermal, siendo esta la principal etapa de mineralización en Candelaria (Williams-Jones y Migdisov, 2014; Del Real et al. 2021). La paragénesis de la etapa de alta temperatura en el distrito Candelaria-Punta del Cobre es idéntica a la paragénesis de mineralización/alteración documentada en los depósitos andinos de IOA como Cerro Negro, El Romeral, Los Colorados y Marcona (Del Real et al., 2021; Rojas, 2018; Chen et al., 2010; Salazar et al., 2020; Reich et al., 2016;

Espinoza et al., 1996; Palma et al., 2021; Childress et al., 2020; Barra et al., 2017; Huang et al., 2019). Por su parte las temperaturas más bajas de actinolita relacionadas con la etapa de IOCG rica en Cu, son consistentes con las condiciones de solubilidad de la calcopirita obtenidas experimentalmente (Del Real et al., 2021; Migdisov et al., 2014; Williams-Jones y Migdisov, 2014). Con evidencias similares presentadas por distintos autores (Espinoza et al., 1996; Knipping et al. 2015a; Arévalo et al., 2006; Palma et al., 2021; Childress et al., 2020; Barra et al., 2017; Huang et al., 2019) los depósitos del distrito Candelaria-Punta del Cobre parecen presentar el mismo patrón de mineralización y enfriamiento de fluidos que el observado en el yacimiento Candelaria.

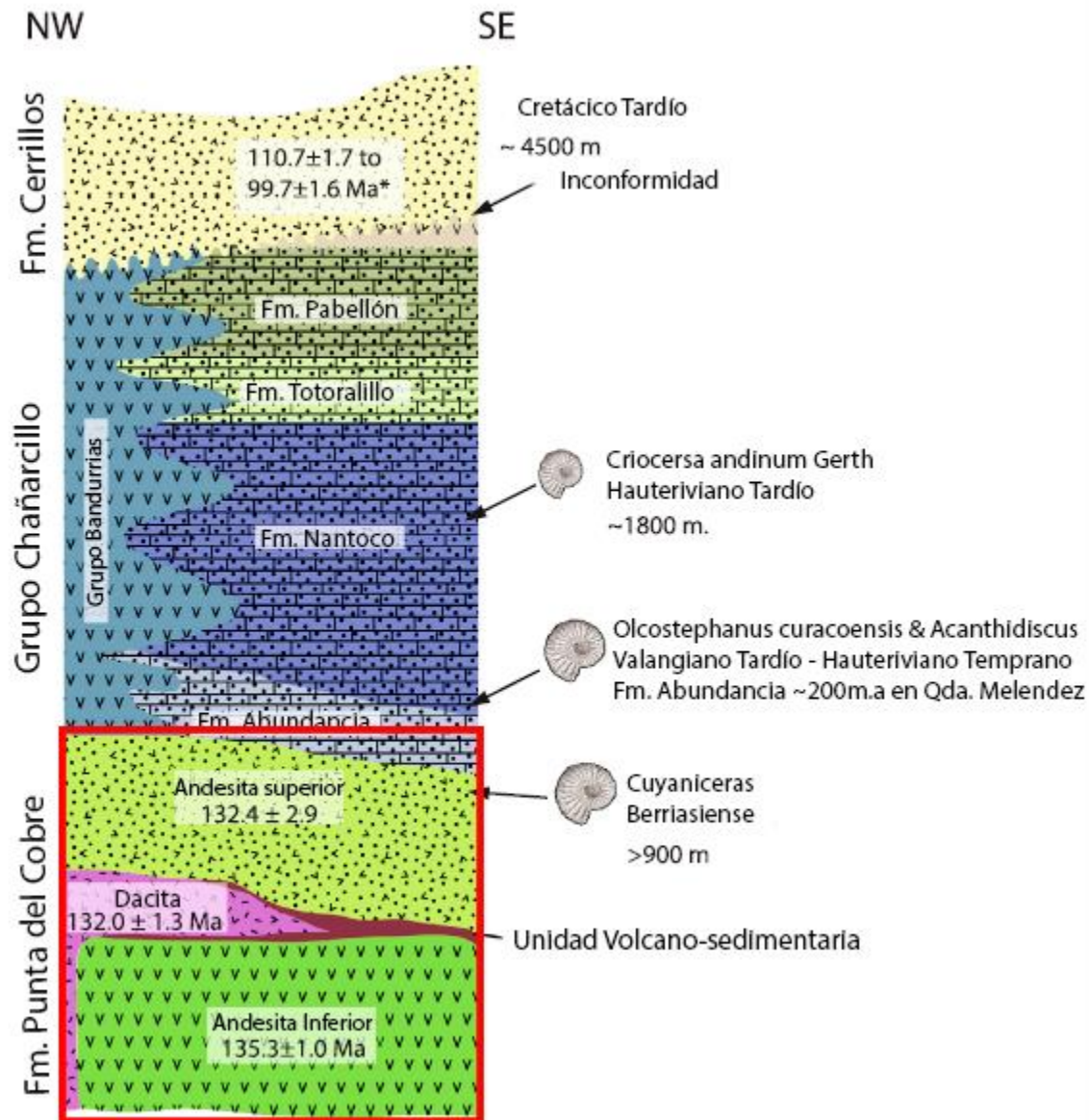


Figura 3.3. Columna estratigráfica del distrito Candelaria-Punta del Cobre con las principales edades geocronológicas y paleontológicas. En rojo se indican las unidades mineralizadas de la Formación Punta del Cobre (extraído de del Real et al., 2020).

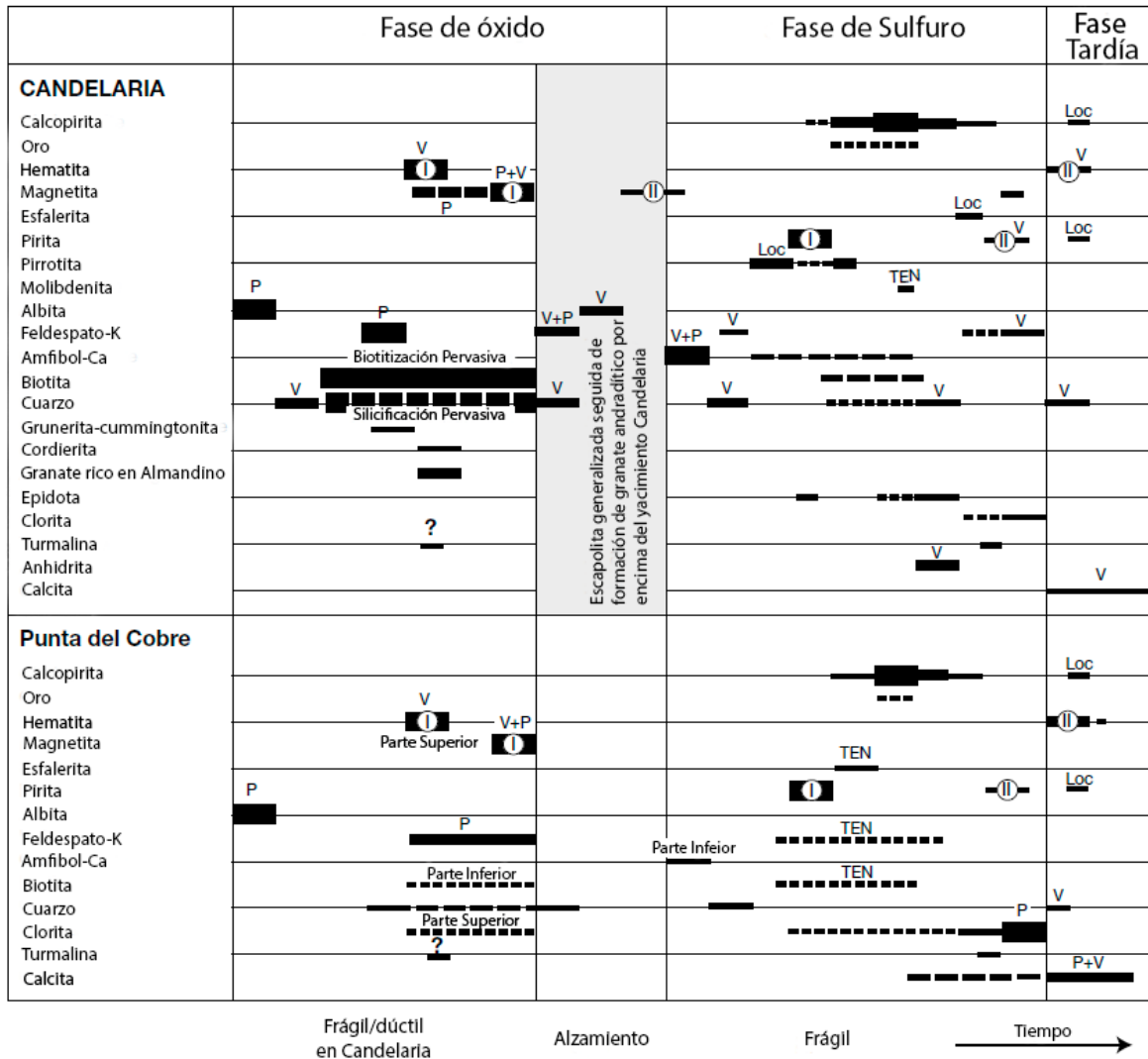


Figura 3.4. Secuencia paragenética de la mineralización de IOCG en el distrito de Candelaria-Punta del Cobre. El grosor de las líneas representa aproximadamente la abundancia relativa del mineral en cuestión, pero no tiene implicaciones cuantitativas respecto a otras especies minerales. Se muestran los eventos de vetillas distinguibles más relevantes en Candelaria. El metamorfismo de contacto máximo en Candelaria, indicado por el campo sombreado en gris, está representado por la amplia alteración sódica de escapolita-piroxeno-y granate por encima del yacimiento. Abreviaturas: P = alteración generalizada, TEN = tentativamente; V = vetillas; ? = incierto (modificado de Marschik y Fontboté, 2001a).

3.2.1.2. Relación entre elementos (mayores y trazas) y eventos mineralizadores

Datos de elementos traza para la magnetita y la pirita junto con los datos de $\delta^{37}\text{Cl}$ del sistema de Candelaria, señalan que los fluidos hidrotermales del distrito Candelaria-Punta del Cobre proceden de magmas silicatados intermedios a máficos, de composición entre gabros y diorita, formados en el arco magmático del Jurásico Superior al Cretácico Inferior del norte

de Chile (Del Real et al., 2020; Rodriguez-Mustafa et al., 2020; Chiaradia et al., 2006; Scheuber y Gonzales, 1999; Pichowiak, 1994).

La caracterización de las variaciones de los elementos traza dentro de la pirita puede servir como monitor de los cambios en la evolución del fluido hidrotermal (Huston et al., 1995; Large et al., 2009; Reich et al., 2013, 2016; Gregory et al., 2015; Tardani et al., 2017 como se citó en del Real et al., 2020). Las variaciones pueden atribuirse a distintos factores como a (1) cambios en la temperatura del fluido hidrotermal y la partición de los elementos traza entre las fases minerales coexistentes a diferentes temperaturas (Abraitis et al., 2004; Keith et al., 2016); (2) cambios en las condiciones de actividad redox y H^+ con la solubilidad de los elementos traza disminuyendo o aumentando a diferentes fO_2 y pH, respectivamente (Thomson et al., 1993; Keith et al., 2017); (3) cambios en la composición del fluido hidrotermal que se reflejarán en la química de la pirita (Huston et al., 1995; Abraitis et al., 2004; Tardani et al., 2017).

La química de la pirita combinada con los análisis in situ de $\delta^{34}S$ proporciona información útil sobre las fuentes de los depósitos de IOCG (Del Real et al., 2018). Las relaciones Co/Ni combinadas con Se/S y $\delta^{34}S$ sugieren que los fluidos responsables de la mineralización principal en el distrito Candelaria-Punta del Cobre tienen un origen magmático-hidrotermal. Los altos valores de $\delta^{34}S$ en la pirita de las vetas tardías sugieren que un fluido derivado externamente entró en el sistema hacia el final del evento mineralizador principal (Del Real et al., 2020).

La principal etapa de mineralización de Cu en el distrito Candelaria-Punta del Cobre habría tenido lugar cuando el fluido o fluidos ascendentes se enfriaron a temperaturas inferiores a $\sim 550^\circ C$ precipitando calcopirita. La evidencia obtenida de los elementos mayores y traza apoyan la influencia de la mineralización es de origen magmático-hidrotermal con una incursión tardía de fluidos derivados de la cuenca (Del Real et al, 2018, 2020, 2021).

3.2.1.3. Edad de mineralización

La información disponible (Arévalo et al., 2006; Marschik y Fontboté, 2001a; Marschik y Söllner, 2006; Mathur et al., 2002; Del Real et al., 2018), indica que los depósitos del distrito Candelaria-Punta del Cobre se formaron en un extenso sistema hidrotermal que duró por lo menos 6 Ma (~ 118 -112 Ma).

La influencia térmica del Plutón San Gregorio promovió la formación de la zona de cizalla de Ojanco, y puede ser precursor del evento mineralizador del distrito (Sillitoe, 2003; Arévalo et al 2006). Datación de biotita verde, vetas de esquistos y brechas sulfurosas (Del Real et al., 2018) indica que la edad de mineralización sería de $\sim 111,7 \pm 0,8$ siendo consistente con la edad de $^{40}\text{Ar}/^{39}\text{Ar}$ reportada por Ulrich y Clark. (1999). Las edades de alteración publicadas por Marschik y Fontboté. (2001a) para el yacimiento de Candelaria oscilan entre 117 y 111 Ma. Estas edades sugieren que la etapa de mineralización principal fue ligeramente más joven, aunque no se descarta la presencia de mineral de pre-deformación como sugieren las edades registradas por Mathur et al (2002) y una edad U/Pb de $124,9 \pm 0,4$ Ma reportada por Pop et al. (2000).

Las paragénesis reportadas por Arévalo et al. (2006) indican que el anfíbol cálcico, es anterior a la fase principal de calcopirita. A su vez, la edad de la molibdenita es consistente con otras edades de biotita y anfíbol de 115 Ma obtenidas tanto en Candelaria como en el distrito de Punta del Cobre (Marschik et al., 1997; Ullrich et al., 1999; Marschik y Fontboté, 2001a).

Marschik et al. (1997), asigna edades radiométricas de ~ 115 Ma para el evento mineralizador principal, la cual estaría relacionada a la actividad tectono-magmática de la Cordillera de la Costa, bajo un régimen sinistral-transtensional del Jurásico Tardío. Arévalo et al. (2006) sugiere que la etapa principal de mineralización fue coetánea al emplazamiento del complejo plutónico de San Gregorio y al desplazamiento en la zona de cizalla Ojancos-La Florida. Del Real et al. (2018) señala que el volcanismo ocurrió entre ~ 135 y 132 Ma, la sedimentación de la cuenca de Chañarcillo se produjo entre ~ 132 - 130 Ma y el emplazamiento del batolito de Copiapó ocurrió entre ~ 118 - 110 Ma. Una serie de diques alterados y mineralizados con la fase principal de mineralización de sulfuro de Cu proporcionan una edad de mineralización mínima de ~ 115 Ma, pero también parecen indicar que la mineralización de la etapa tardía continuó hasta al menos 112 Ma. Esto es consistente con las relaciones observadas en la zona de cizalla de Ojancos, que sugieren que la mineralización ocurrió tanto antes como después del emplazamiento ~ 115 Ma de la fase San Gregorio del batolito de Copiapó.

La mineralización en el distrito Candelaria-Punta del Cobre está alojada en rocas que son 15 Ma más antiguas que el evento principal de Cu (Del Real et al., 2018); por tanto, para que

los fluidos hidrotermales asciendan las fallas sinistralas habrían sido cruciales para el flujo de fluidos hacia la estratigrafía volcano-sedimentaria más antigua.

3.2.2. Distrito Mantoverde

El distrito Mantoverde está ubicado en la Cordillera de la Costa entre los 26°25'-26°45'S (Figura 3.1 y 3.5). El distrito se encuentra emplazado en rocas volcánicas e intrusivas del Jurásico – Cretácico Temprano (Benavides, 2007). Las rocas volcánicas donde se hospeda principalmente la mineralización en Mantoverde se han asignado a las Formaciones La Negra y Punta del Cobre (Benavides, 2007). Fallas subsidiarias del SFA de tendencia NNW a NW conectadas por la Falla Mantoverde juegan un rol importante en la mineralización de IOCG del distrito Mantoverde, siendo los principales conductos de los fluidos magmáticos-hidrotermales en las rocas ígneas menos permeables generando un intenso veteado en dichas rocas (Marschik y Fontboté 2001; Benavides et al., 2007; Rieger et al. 2010a).

3.2.2.1. Mineralización y alteración

La mineralización principal en el sector de Mantoverde se extiende por unos 10 km en dirección N-S y consiste en brechas y stockwork de calcopirita y hematita especular, cuerpos con magnetita-apatito $\pm$ pirita; y se estima que se formó hace ~ 119 Ma (Vila et al., 1996). Diversos autores señalan que distintos agentes podrían haber incidido en la mineralización del distrito tales como fluidos acuosos juveniles, las aguas de formaciones sedimentarias, aguas meteóricas, el agua de mar y/o las salmueras evaporíticas (Scheuber y Andriessen 1990; Marschik et al. 2003; Marschik y Söllner 2006).

Según el modelo paragénético de Benavides et al. (2007), el distrito Mantoverde incorpora cuatro etapas. La etapa I está dominada por un metasomatismo generalizado de potasio y hierro que convirtió las rocas granitoides y volcánicas en ortoclasa y magnetita, respectivamente. La etapa II comprende la alteración clorítica y sericítica, y la formación de vetas. En la etapa III ocurrió la deposición de calcopirita, alojada en vetas de calcita y en brechas hidrotermales y stockworks dominados por hematita especular. Durante la etapa IV se registran enjambres de vetas de calcita-cuarzo de origen hidrotermal. Las relaciones de fraccionamiento de isótopos estables y la microtermometría de inclusión de fluidos definen una evolución térmica retrógrada, desde temperaturas $>460^{\circ}\text{C}$ en la etapa I, pasando por

~350°C en la etapa II, hasta ~210° a 280°C en la etapa III del mineral, y ~110° a 240°C en la etapa IV.

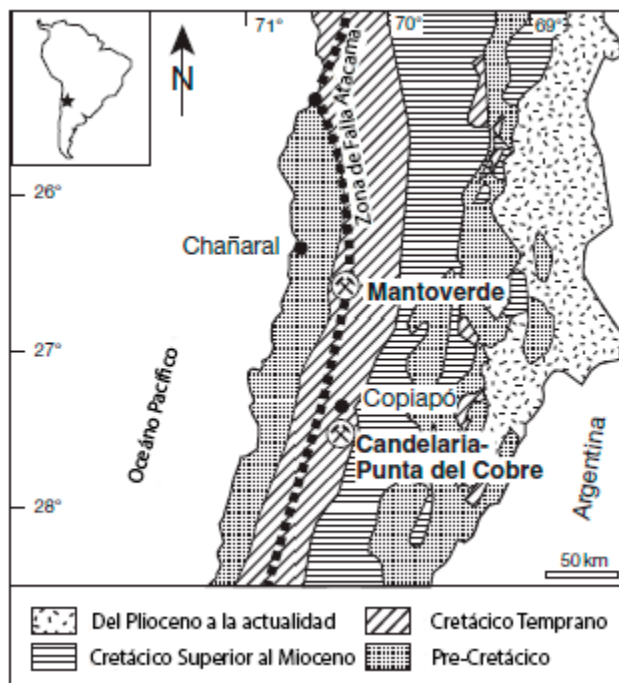


Figura 3.5. Mapa geológico esquemático del área alrededor de Copiapó y Chañaral, norte de Chile. La localización de los dos distritos mayores de IOCG, Mantoverde y Candelaria-Punta del Cobre, son indicador. ⊗ = Mina (extraído de Rieger et al., 2010a).

La mineralogía de sulfuros hipógena es simple y consiste en calcopirita y pirita. La bornita y la digenita que sustituyen a la calcopirita están presentes sólo en la base de la zona oxidación (Rieger et al., 2012). Dentro del distrito se ha reconocido una marcada zonación en la distribución de los óxidos de hierro (Rieger et al. 2010a; Williams et al. 2005; Zamora y Castillo 2001). Las zonas dominadas por la especularita ocurren en niveles poco profundos en la parte norte del distrito, a diferencia de los minerales dominados por magnetita que están presentes en la parte sur del distrito, así como en niveles más profundos de la parte norte, por debajo de las zonas especulares (Rieger et al., 2012).

A medida que las temperaturas descienden y la química de los fluidos evolucionaba a través de reacciones fluido-roca, la formación de conjuntos de alteración K-Fe, brechas y sulfuros dio lugar a la mineralización de IOCG del grupo de la magnetita y la hematita. La mineralización del distrito Mantoverde se subdivide en tres etapas hidrotermales (Marschik et al. 2003a; Rieger et al. 2010a, 2012) (Figura 3.6):

1. Etapa temprana de óxido de hierro con temperaturas $>400^{\circ}\text{C}$ caracterizada por un intenso metasomatismo de hierro que formó hematita, la cual posteriormente es reemplazada y cubierta por magnetita en profundidad, mientras que en niveles poco profundos la hematita permaneció inalterada (similar a la Etapa I mencionada en el modelo paragénético de Benavides et al. (2007)).
2. Etapa de sulfuros a $\sim 350\text{-}250^{\circ}\text{C}$, la que tuvo lugar la principal mineralización de Cu y oro (equivalente a las Etapas II y III propuestas por Benavides et al. (2007)).
3. Etapa hidrotermal tardía $<250^{\circ}\text{C}$, que contribuyó localmente con cantidades menores o trazas de hematita y sulfuros, y abundante calcita (Figura 3.6) (similar a la Etapa IV de Benavides et al. (2007)).

En general, ambos modelos describen etapas de mineralización similares, aunque con diferentes nomenclaturas.

La alteración del distrito Mantoverde se caracteriza por un intenso metasomatismo potásico, silicificación y carbonitización, y carece del intenso componente de Na y/o Ca que se observa en el distrito de Candelaria-Punta del Cobre (Marschik y Fontboté 2001a). La alteración relacionada con la mineralización de Cu se caracteriza por conjuntos que contienen cantidades variables de feldespato K, clorita, sericita, cuarzo y/o carbonatos (Rieger et al., 2012; Marschik et al. 2003). Las rocas más intensamente alteradas se transformaron en agregados granulares de grano fino y color rosa (López et al., 2014).

La formación de óxido de hierro estuvo acompañada por una alteración generalizada del feldespato K, vetas de cuarzo y una silicificación generalizada. La alteración por clorita y sericita marca el final de la etapa de óxido de hierro. La principal mineralización de Cu-Au estaba asociada a vetas de $\text{cuarzo} \pm \text{K-feldespato}$, y a vetas de $\text{sericita} \pm \text{cuarzo}$. La carbonatización generalizada y el intenso veteado de $\text{calcita} \pm \text{especularita}$ son características típicas de la etapa hidrotermal tardía, que contribuyó localmente con cantidades menores o trazas de pirita y calcopirita (Rieger et al. 2010a).

3.2.2.2. Edad de mineralización

Se han publicado varios estudios geocronológicos de alteración y mineralización para Mantoverde (Vila et al., 1996; Mathur et al., 2002; Gelcich et al., 2005; Barra et al., 2017). Un estrecho vínculo temporal entre la mineralización y el emplazamiento del plutón en

Mantoverde fue inferido a partir de la datación de U-Pb de la titanita hidrotermal de uno de los diques félsicos intrusivos que muestra alteración potásica ($126,4 \pm 0,5$ Ma; Gelcich et al., 2003). La edad del Cretácico Temprano de la mineralización está restringida por la geocronología K-Ar en sericita hidrotermal que arroja 121 ± 3 y 117 ± 3 Ma (Vila et al. 1996) y una edad isócrona de Re-Os derivada de la magnetita que da 116 Ma (Mathur et al. 2002). Barra et al. (2017) informaron de tres dataciones Re-Os que abarcan un intervalo de $148 \pm 2,2$ a $115 \pm 1,7$ Ma para la mineralización de pirita y magnetita + pirita en Mantoverde.

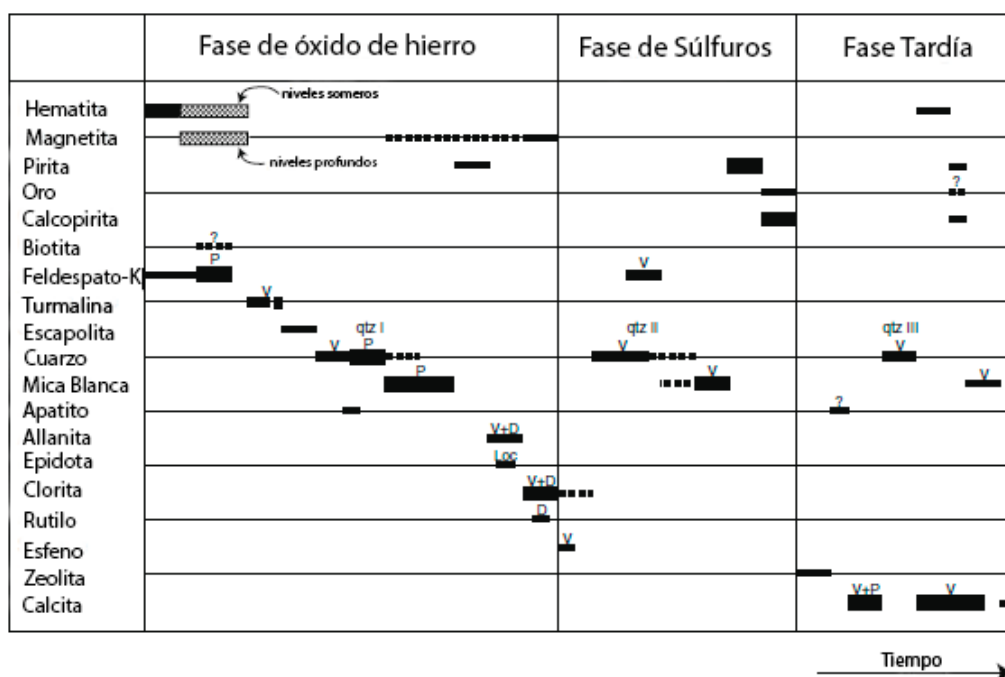


Figura 3.6. Secuencia paragenética de la mineralización IOCG en el distrito Mantoverde. *D* diseminaciones, *Loc* local, *P* alteración generalizada, *qtz* cuarzo, *V* vetas. (extraído de Rieger et al., 2012).

Verdugo-Ihl et al. (2022) realizaron dataciones de hematita U-Pb para el yacimiento de Mantoverde. Los datos de LA-ICP-MS normalizados utilizando material de referencia de hematita desarrollado recientemente, arrojan dataciones de $105,61 \pm 6,71/14/19,39$ Ma en Mantoverde. La hematita es rica en Ti, pero carece de exsoluciones de ilmenita. Esto implica temperaturas de cristalización mínimas de $\sim 500^\circ\text{C}$, por encima del solvente del sistema $\text{FeTiO}_3\text{-Fe}_2\text{O}_3$, apoyando una fuente de fluido de origen magmático. La nueva fecha para Mantoverde, unida a los datos publicados anteriormente, podría indicar la superposición de varios eventos mineralizadores (Verdugo-Ihl et al., 2022; Childress et al., 2020).

Los sistemas de IOCG en los distritos de Mantoverde y Candelaria- Punta del Cobre están zonificados: el cobre y el oro están asociados a la hematita en niveles poco profundos o en lugares distales con respecto a los conductos de fluidos inferidos, y la magnetita aparece en profundidad o en las partes proximales del sistema hidrotermal (Marschik y Fontboté 2001a; Rieger et al. 2010b).

3.2.2.3. Relación entre elementos (mayores y trazas) y eventos mineralizadores

Los datos isotópicos de Fe y O, datos de elementos trazas, temperaturas de homogenización de las inclusiones fluidas para la magnetita de la etapa de óxido de hierro y la hematita de la etapa tardía del distrito de Mantoverde (Childress et al., 2020); combinados con los datos publicados previamente (Benavides et al. 2007; Rieger et al. 2010a,b, 2012) apoyan una génesis magmática-hidrotermal de enfriamiento que formó la etapa de óxido de hierro rica en magnetita a temperaturas entre 530°C y 278°C, seguida por la mineralización de la etapa de sulfuro de nivel intermedio a superficial sobre el rango de 325°C a 221 °C, y la etapa de hematita y calcita de nivel tardío sobre el rango de 462°C y 221 °C.

Los núcleos de magnetita están enriquecidos en Ti, V, Mn, Si, Al, Mg y Ca en relación con los bordes circundantes, lo es consistente con el origen de la magnetita de un fluido inicialmente caliente y gradualmente enfriado (Childress et al., 2020). La tendencia al enfriamiento está respaldada por los datos de microtermometría de las inclusiones fluidas con magnetita en el cuarzo de la etapa de óxido de hierro de Mantoverde, lo que indica que la mineralización se produjo en un rango de temperatura de 530°C a 278°C (Rieger et al., 2012; Childress et al., 2020).

CAPÍTULO 4: METODOLOGÍA

4.1. Descripción general de la metodología

En el presente capítulo se abordan los procedimientos, materiales y/o herramientas necesarias para la ejecución de esta tesis. Para el desarrollo del modelo de ML se usó herramientas de código abierto. Se destaca el uso del lenguaje de programación *Python* en su versión 3, sus librerías, cuadernos de texto y código (notebooks) como *Jupyter*. En el diagrama de trabajo (Figura 4.1.) se ilustran las etapas sucesivas de recopilación de datos, análisis exploratorio, preprocesamiento, agrupación, clasificación, validación e interpretación.

En la primera etapa se recolectó información bibliográfica disponible sobre la génesis, características y distribución de los depósitos IOCG e IOA, además de información disponible del área de estudio, ya sean estudios geológicos y metalogénicos distritales. Adicionalmente se revisaron artículos y estudios vinculados al uso de modelos de ML en el estudio y exploración de depósitos minerales en general y REE.

Con el fin de corroborar y validar el correcto funcionamiento de los algoritmos, se utilizó el código proporcionado por Cartagena (2021) con su correspondiente conjunto de datos geoquímicos. Para la elaboración del modelo de ML se creó una base de datos con los datos del área de estudio.

La base de datos consiste en una recopilación de datos geoquímicos previamente publicados por diversos autores, los cuales están distribuidos en los distritos Candelaria-Punta del Cobre y Mantoverde (Figura 4.2). Posteriormente los datos recopilados fueron sometidos a un control de calidad para homogeneizarlos. Mediante la plataforma web Kaggle¹ se buscaron modelos de ML y conjuntos de datos para validar el funcionamiento de los algoritmos escogidos. Una vez validado el funcionamiento de los algoritmos se aplicaron los algoritmos. Posteriormente, se procedió a evaluar el rendimiento de los algoritmos y analizar los resultados obtenidos. Finalmente, se interpretaron los resultados obtenidos y sus implicancias geológicas.

¹ www.kaggle.com

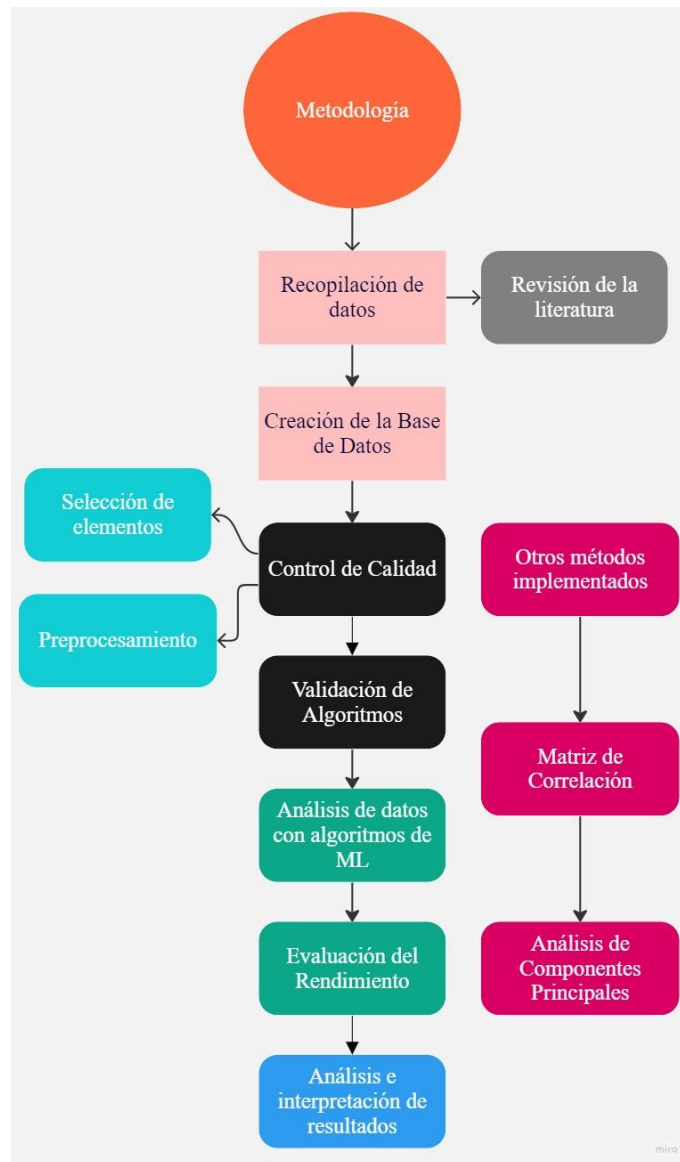


Figura 4.1. Diagrama que muestra el flujo de trabajo seguido para la realización de este trabajo.

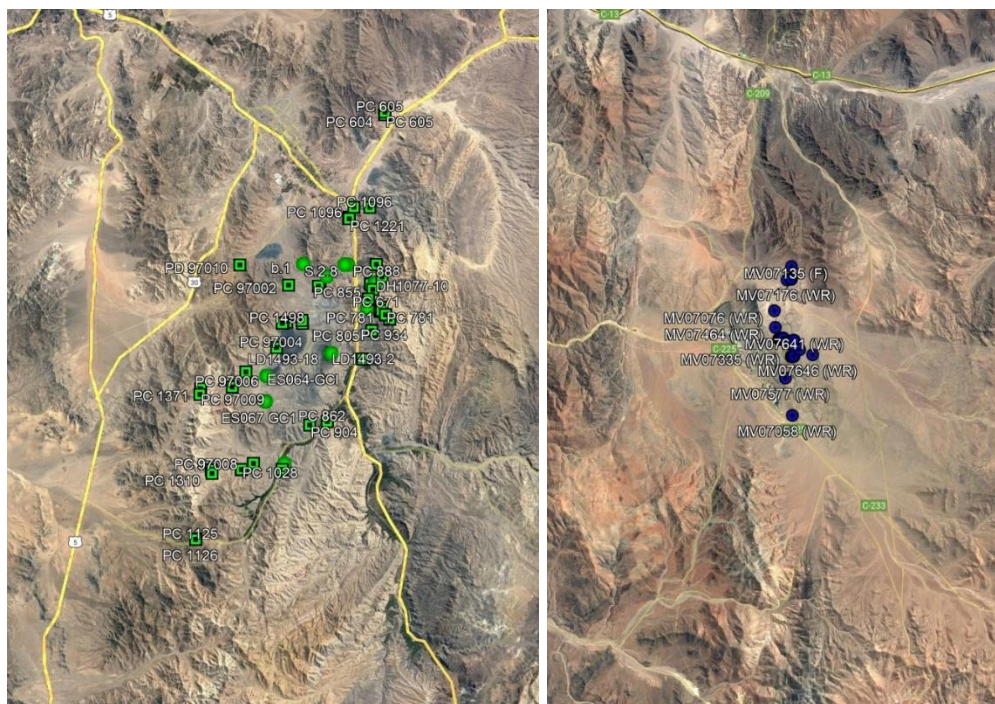


Figura 4.2. Distribución de las muestras recopiladas de los distritos Candelaria-Punta del Cobre (izquierda) y Mantoverde (derecha).

4.2. Herramientas

Antes de adentrarse en el procedimiento de creación del modelo de ML, se describen las herramientas usadas para la elaboración de esta tesis.

4.2.1. Anaconda

Anaconda es un software libre que proporciona un conjunto de herramientas de los lenguajes de *Python* y *R*, orientada a dar soporte integral al desarrollo científico (Rondón, 2020). Anaconda simplifica la gestión e implementación de paquetes (forma de organizar grupos de clases) (Pythonista, 2020). Además, da acceso a diferentes entornos de codificación. Estos entornos también conocidos como Entornos de Desarrollo Integrados (IDE, por sus siglas en inglés), son plataformas o aplicaciones que facilitan enormemente el desarrollo de código (Rolon-Mérette et al., 2016). Para esta tesis se usó *Anaconda* debido a que habilita el entorno de trabajo con *Jupyter Notebook*.

4.2.2. Lenguaje de programación *Python*

Python es un lenguaje de alto nivel de programación, no tipado, dinámico y multiparadigma cuya filosofía es su fácil sintaxis (Álvarez, 2021; Bravo, 2019; Venners, 2003). Es

ampliamente utilizado para aplicaciones web, el desarrollo de software, la ciencia de datos, el ML entre otros (Bravo, 2019; Venners, 2003). Posee una licencia de código abierto, lo que significa que es de libre uso ya sean para propósitos académicos, comerciales y/o personales (Thanaki, 2017). Asimismo, *Python* posee un gran número de librerías que están o pueden ser implementadas lo que favorece a un rápido desarrollo de aplicaciones, siendo las más conocidas las relacionadas con el ámbito de inteligencia artificial y manejo de operaciones matemáticas (Thanaki, 2017; Bravo, 2019). Debido a sus características anteriormente señaladas, se ha trabajado con este lenguaje para esta tesis.

4.2.3. Librerías

4.2.3.1. Numpy

NumPy es un paquete fundamental para la computación científica con *Python* especializada en el cálculo numérico y análisis de datos (Alberca, 2022a). *NumPy* se especializa en el procesamiento numérico a través de matrices (*arrays*), donde las matrices permiten realizar operaciones elemento a elemento (Bressert, 2012).

4.2.3.2. Pandas

La librería *Pandas* proporciona manejo flexible y análisis de estructuras de datos que facilitan la manipulación y tratamiento de dichos datos (Alberca, 2022b). Es usado para el trabajo con *dataframe* (estructura de datos tabulares bidimensionales de tamaño variable y potencialmente heterogéneo) (McKinney, 2011). Además, permite realizar análisis exploratorio de datos, así como también, gráficos estadísticos (Bravo, 2019).

4.2.3.3. Matplotlib

Matplotlib es una librería de *Python* para el diseño de gráficos 2D de calidad (Ari y Ustazhanov, 2014; Tosi, 2009). Está destinado principalmente a la visualización de datos científicos, de ingeniería y financieros (Barrett et al., 2005). Este escrito para el uso de *NumPy* y algún otro código de extensión. Con pocas líneas de código, genera gráficos, histogramas, espectros de potencia, gráficos de barras, gráficos de error, gráficos de dispersión, entre otros (Ari y Ustazhanov, 2014).

4.2.3.4. Scikit-learn

Debido a que *Python* es lenguaje de código abierto, incontables usuarios han podido diseñar un número demasiado elevado de librerías en donde se puede encontrar técnicamente todas las técnicas de ML. Sin embargo, resulta en un problema debido a que cada librería de ML tiene una sintaxis, estructura e implementación propia, dificultando su uso (Davis, 2021). En este contexto se ejecuta *Scikit-learn*, el cual es una librería de código abierto que unifica bajo un único marco los principales algoritmos y funciones, facilitando en gran medida todas las etapas de preprocesado, entrenamiento, optimización y validación de modelos predictivos; tanto para problemas de aprendizaje supervisado como no supervisados (Hao y Ho, 2019; Amat, 2020; Bravo, 2019).

4.2.3.5. Jupyter Notebook

Jupyter Notebook es un entorno de ejecución informático interactivo de código abierto, basado en la web que permite crear y compartir documentos y códigos (Project Jupyter, 2016; Ionos, 2019).

4.3. Base de Datos

En este estudio el conjunto de datos C-E es una colección de 91 muestras geoquímicas de rocas de los distritos Candalaria-Punta del Cobre y Mantoverde previamente publicadas por diversos autores (Anexo 1). Para cada muestra (fila) se tienen 34 características (columnas), las que incluyen concentraciones de elementos mayores y REE. Debido a las distintas procedencias de las rocas, la clasificación litológica presenta diferentes nomenclaturas y criterio. Para homogeneizar la clasificación litológica, debido a que la litología es una variable categórica, cada muestra de los datos compilados se les asignó un valor numérico siguiendo al tipo de litología correspondiente. Como conjunto de entrada de los datos del área de estudio se seleccionaron 29 características las cuales corresponden a óxidos mayores y REE. Se excluyeron las variables como el ID de muestra, referencia, coordenadas este y oeste, y método analítico debido a que no aportan información relevante para este análisis.

4.3.1. Control de Calidad

Debido a que la base de datos C-E se compiló a partir de diversos trabajos, en los cuales a su vez las concentraciones geoquímicas fueron obtenidas utilizando diferentes métodos de análisis, por ende, se tienen datos ruidosos, ausentes o confusos, por lo cual se requirió un

control de calidad sobre los mismos. Este control de calidad comprendió limpieza de datos, normalización, estandarización y detección de datos atípicos. El objetivo del proceso de control de calidad y homogeneización consiste en garantizar que los datos geoquímicos de los distintos trabajos efectivamente representen muestras asociadas al proceso de mineralización y formación del depósito, y no a otros factores (p.ej. contaminación de la muestra, eventos de meteorización, etc.). Producto de que no todas las muestras tienen las mismas columnas de datos, se seleccionaron ciertas características. Se eligieron las características que contenían registros más completos en los conjuntos de datos. Las características seleccionadas fueron: tipo de roca, SiO₂, Al₂O₃, Fe₂O₃, MgO, CaO, Na₂O, K₂O, TiO₂, P₂O₅, MnO, Sr, Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Dy, Ho, Er, Tm, Yb y Lu.

4.4. Validación de Algoritmos

Con el objetivo de corroborar el funcionamiento correcto de los algoritmos seleccionados, se replicaron las líneas de código indicadas por Cartagena (2021)² para *K-means* y agrupación jerárquica, con su respectiva base de datos geoquímica, además de emplear los tutoriales de la librería *Scikit Learn* disponibles en su página web³. A partir de la replicación del código se obtuvo un resultado idéntico al obtenido por Cartagena (2021), obteniendo con *K-means* la misma cantidad de clústeres (5) y la misma asignación de cada dato para su respectivo clúster, aunque cabe señalar que existieron pequeñas discrepancias debido al número de iteraciones efectuadas (se realizaron un total de 300 iteraciones). Al momento de realizar la estandarización de los datos, los valores obtenidos tienen una diferencia del cuarto decimal de los valores obtenidos del modelo original. En el caso de la aplicación del algoritmo de agrupación jerárquica, se obtuvo las mismas correlaciones jerárquicas entre las muestras y la misma cantidad de clústeres (4). Para la implementación de PCA se siguió el procedimiento expuesto por Amat (2020). Una vez validado el funcionamiento de los algoritmos se procedió a aplicarlos en la base de datos geoquímica.

4.5 Análisis de datos con algoritmos de *Machine Learning*

La Figura 4.3 muestra el flujo de trabajo seguido para analizar cada base de datos. Para aplicar los algoritmos de ML primero se importó la base de datos de trabajo al entorno de trabajo

² Cartagena, 2021. Recuperado el 02 de noviembre de 2022 de [cluster baseline | Kaggle](#)

³ <https://scikit-learn.org/stable/>

Jupyter. Posteriormente se aplicaron sobre dicha base de datos los algoritmos de ML no supervisados. Luego se evaluó el rendimiento de cada algoritmo aplicando 11 métricas. Finalmente, se interpretaron los resultados obtenidos. En el Anexo 2 se incluye el enlace con el código en formato *Jupyter* que se utilizó.

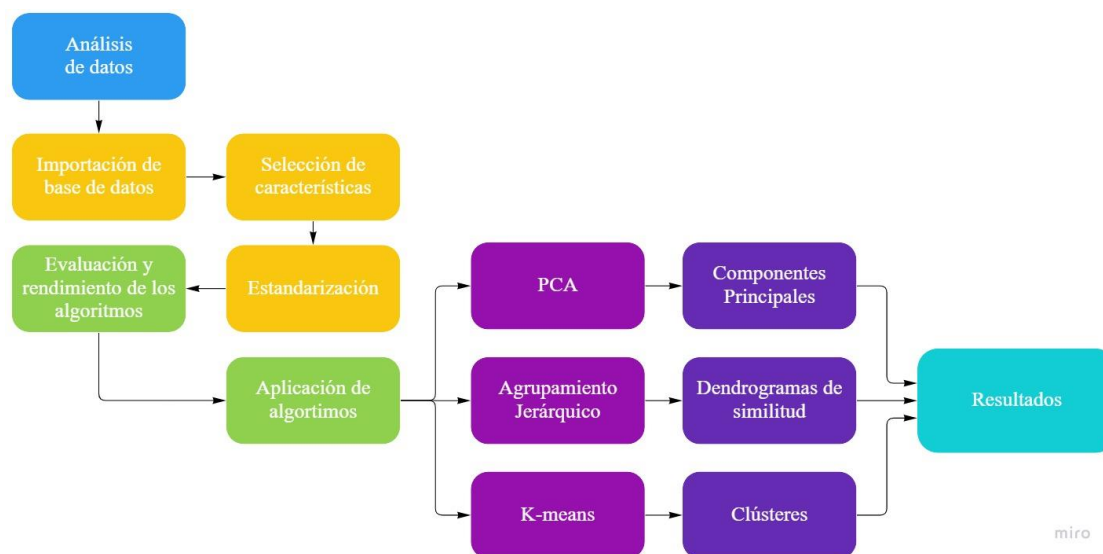


Figura 4.3. Diagrama de Flujo de trabajo del análisis de datos con algoritmos de *machine learning*.

4.5.1 Preprocesamiento

Una vez importadas las bases de datos al entorno *Jupyter*, se eliminaron las columnas de datos que no se usaron durante la creación del modelo (p.ej. depósito, profundidad, SUM, y LOI). Luego, se verificó que el conjunto de datos no contenga casillas de información vacías, (estas casillas aparecerán con la nomenclatura NaN). Debido a que la concentración de elementos químicos depende de varios factores, es difícil tratar de rellenar las casillas vacías con el promedio estadístico o mediana de la columna correspondiente, lo cual podría generar datos falsos y/o sesgados. Es por ello por lo que se decidió aplicar la función *dropna* de la librería *Pandas*, la cual elimina las columnas o filas (según convenga) que contengan casillas con datos faltantes.

La base de datos presenta 30 celdas sin datos, por lo que se aplicó la función *dropna* y se eliminaron 10 muestras, quedando 81 muestras para su análisis. Posteriormente los datos geoquímicos fueron normalizados mediante la transformación logarítmica de los datos, propuesta por Goovaerts (1997). Luego de la selección de columnas y muestras se procedió

a la estandarización de valores, debido a que los datos de los conjuntos son de distintos tipos y escalas (p.ej. porcentajes, decimales, enteros o valores superiores a cien). La estandarización de datos facilita que los datos de origen sean internamente coherentes, en otras palabras, que cada tipo de datos tenga el mismo tipo de contenido y de formato (Choi, 2018). Para ello se utilizó la función *StandardScaler* de la librería *Sklearn*. La función *StandardScaler* transforma los datos de manera que su distribución tenga un valor medio de cero y una desviación estándar de 1. Dada la distribución de los datos, cada valor en el conjunto de datos tendrá restado el valor medio de la muestra, y luego dividido por la desviación estándar de todo el conjunto de datos (Choi, 2018).

4.5.2 Análisis exploratorio de la base de datos

Se realizó un análisis multivariable con el fin de entender la relación entre las variables dependientes y las variables independientes. Finalmente se comprobó si los datos cumplen con los supuestos exigidos por la mayoría de las técnicas multivariantes. Buccianti y Grunsky (2014), Hair et al. (2014) y Dmitrijeva et al. (2019b) han expuesto las dificultades asociadas para realizar análisis de datos geoquímicos, los cuales tienen propiedades matemáticas específicas, que solo contienen información relativa y que están ligadas a sumar valores numéricos como 100% del peso o 1.000.000 ppm. Es por ello por lo que se sigue su recomendación de emplear un enfoque multivariante y la transformación log-ratio de Aitchison (1986), la cual convierte los datos en un espacio real numérico, al cual se pueden aplicar métodos estadísticos estándar sin que se identifiquen correlaciones adulteradas.

Para investigar la relación entre todos los componentes de los datos geoquímicos, se obtuvo una matriz de correlación, donde cada entrada de la matriz es la varianza de la relación logarítmica del par de elementos (Aitchison, 1986). La matriz de correlación ha sido efectiva en hallar correlaciones entre variables en estudios de exploración geoquímica, geología marina, etc. (Zaremotlagh et al., 2016; Dmitrijeva et al., 2019; Lemenkova, 2019). Una matriz de correlación es una tabla que indica los coeficientes de correlación entre varias variables cuantitativas (Figura 4.4). La matriz de correlación fue efectuada con los datos sin estandarizar debido a las propiedades geométricas de las composiciones, y al ser preprocesadas existe la alta posibilidad de pérdida de información en la correlación y/o covarianza.

Variables	X	Y	Z
X	1	0.3	0,75
Y	0,3	1	0,95
Z	0,75	0,95	1

Figura 4.4. Ejemplo de una matriz de correlación. En donde los valores cercanos a uno indican que tienen una relación directa. Por el contrario, valores menores a cero indican relación indirecta. Valores cercanos a cero indican correlación débil o nula.

4.6 Análisis de Componentes Principales

Se empleó el método de Análisis de Componentes Principales (PCA, por sus siglas en inglés) con el objetivo de disminuir la dimensionalidad de la base de datos y encontrar las direcciones con mayor varianza. Para llevar a cabo el PCA, se siguió un proceso estándar en el que se calcularon las componentes principales y se evaluó su contribución a la varianza total de los datos. Las componentes principales se calcularon mediante una transformación lineal de los datos originales, que se centraron en la media y se escalonaron para que tuvieran una varianza unitaria. Luego, se calcularon los valores propios (*eigenvalues*) de la matriz de covarianza de los datos, y se eligieron las componentes principales que explicaron la mayor parte de la varianza total de los datos. Este proceso permitió identificar patrones y relaciones entre las variables y facilitó el análisis de la base de datos geoquímicas con datos de tierras raras.

La figura 4.5 muestra un ejemplo donde la mayor parte de la varianza (44%) se explica con el primer componente principal, el segundo aporta con un poco más del 20%, mientras que el tercero y el cuarto aportan cerca del 10% cada uno. En total, los cuatro componentes principales suman alrededor del 83% de la varianza, lo que se acerca al valor total de 90% observado en la figura 4.5. Para visualizar los resultados del PCA, se utilizó un mapa de calor en el que se graficó la correlación entre las variables y cada una de las componentes principales (ver ejemplo en Figura 4.6). Esto permitió identificar patrones y relaciones entre las variables y las componentes principales, lo que facilitó la comprensión de la estructura de la base de datos geoquímicas con datos de tierras raras y las relaciones entre las variables.

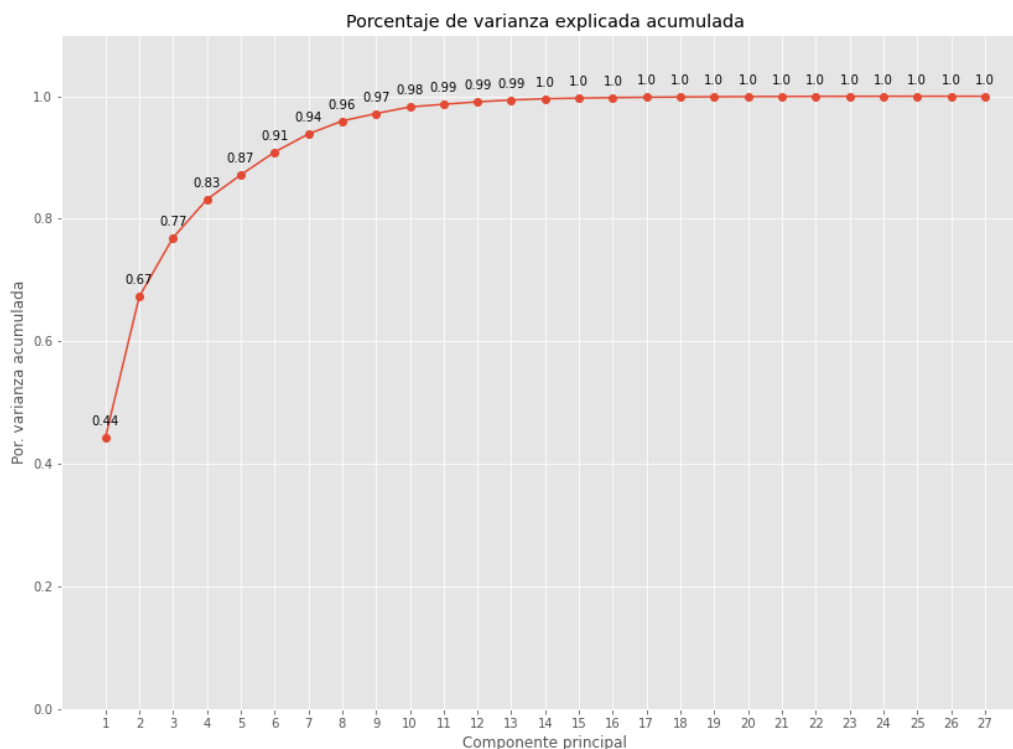


Figura 4.5. Ejemplo del porcentaje de la varianza explicado por cada componente principal resultante de un análisis PCA.

4.7. Algoritmos utilizados y su validación

A continuación, se describe la aplicación de cada algoritmo de ML, así como sus métricas de evaluación de rendimiento. En este trabajo se aplicaron 2 algoritmos de aprendizaje no supervisado: agrupamiento jerárquico y *K-means*. En conjunto a los algoritmos anteriormente mencionados se empleará PCA para facilitar la evaluación visual. Los diferentes pasos se ilustran con figuras obtenidas luego de aplicar los algoritmos sobre el conjunto de datos de ejemplo.

4.7.1 Agrupamiento jerárquico

En el agrupamiento jerárquico se debe definir las medidas de distancias o medidas de similitud que serán implementadas para calcular la vinculación. Para este caso se usa la distancia euclidiana, la cual es la distancia ordinaria en línea recta entre 2 puntos en el espacio euclidiano, con la cual el espacio euclidiano pasa a ser un espacio métrico. El eje vertical del dendrograma es una medida de similitud entre los elementos la cual está directamente relacionada con la matriz de correlación (Murtagh y Contreras, 2012). La Figura 4.7 ilustra un ejemplo de un dendrograma a partir de la similitud de los elementos geoquímicos. Se

puede observar que se formaron 3 clústeres, de igual manera, la Figura 4.8 es un ejemplo de dendrograma que muestra la similitud de las muestras para una misma base de datos, se aprecia que también se generan 3 clústeres potenciales. La altura de las ramas representa la similitud entre los individuos o los clústeres. Mientras más pequeña sea la altura mayor es la cercanía.

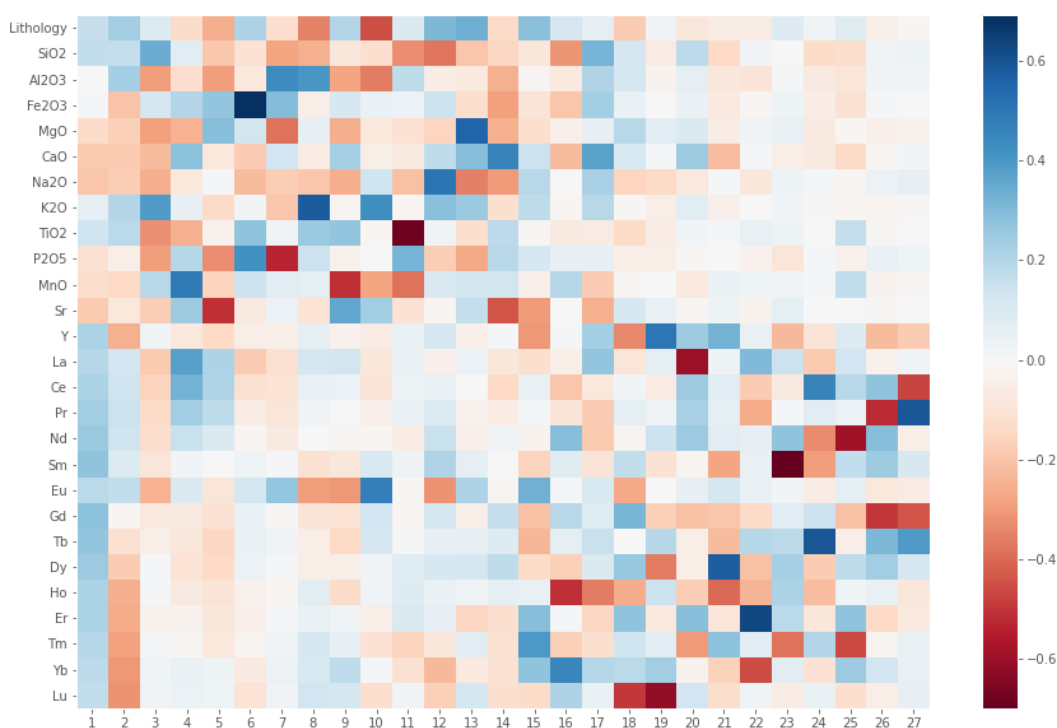


Figura 4.6. Ejemplo de mapa de calor que muestra la influencia de las variables en cada componente principal obtenida. El eje vertical corresponde a las variables. El eje horizontal corresponde a los componentes principales.

Para determinar el número óptimo de clústeres en el Agrupamiento Jerárquico se emplearon 3 métricas para distintos números de clústeres y se seleccionó el número óptimo de clústeres que maximizó estas métricas simultáneamente. La métrica de Calinski-Harabasz se calculó mediante la relación entre la dispersión intra-cluster y la dispersión inter-cluster. Un valor más alto de esta métrica indica una mayor separación entre los clústeres, lo que sugiere un número óptimo de clústeres. El valor de la métrica se calculó para diferentes valores de k (número de clústeres), y se eligió el valor de k que maximizó la métrica de Calinski-Harabasz. La métrica Cophenética se utilizó para evaluar la coherencia de la estructura de *clustering*. Se construyó un dendrograma utilizando un método jerárquico aglomerativo y se calculó el

coeficiente de correlación de Cophenetic entre la matriz de similitud original y la matriz Cophenética. Luego, se repitió este proceso para diferentes valores de k y se eligió el valor de k que maximizó el coeficiente de correlación de Cophenetic. Por último, se utilizó el índice Davies-Bouldin para evaluar la calidad de los clústeres. Esta métrica compara la dispersión dentro de cada clúster con la dispersión entre clústeres, y un valor más bajo indica una mejor calidad de los clústeres. Se calculó el índice para diferentes valores de k , y se eligió el valor de k que minimizó el índice de Davies-Bouldin.

Las métricas, Calinski-Harabasz, el índice Davies Bouldin y Cophenética muestran que en el ejemplo un número óptimo de clústeres es 3. De esta manera, se aseguró una partición de los datos que optimizara la separación entre los clústeres y la cohesión dentro de ellos. La Tabla 4.1 muestra un ejemplo de los valores obtenidos para distintos números de clústeres.

Tabla 4.1. Métricas de evaluación de *clustering* para Agrupamiento Jerárquico.

N° Clústeres	CH	DB	Cophenética
3	15,62	1,15	0.74
2	15,44	1,09	0.74
4	14,35	1.09	0.68
5	13.08	0.81	0.73

Abreviaciones: Calinski-Harabasz (CH), Davies-Bouldin (DB).

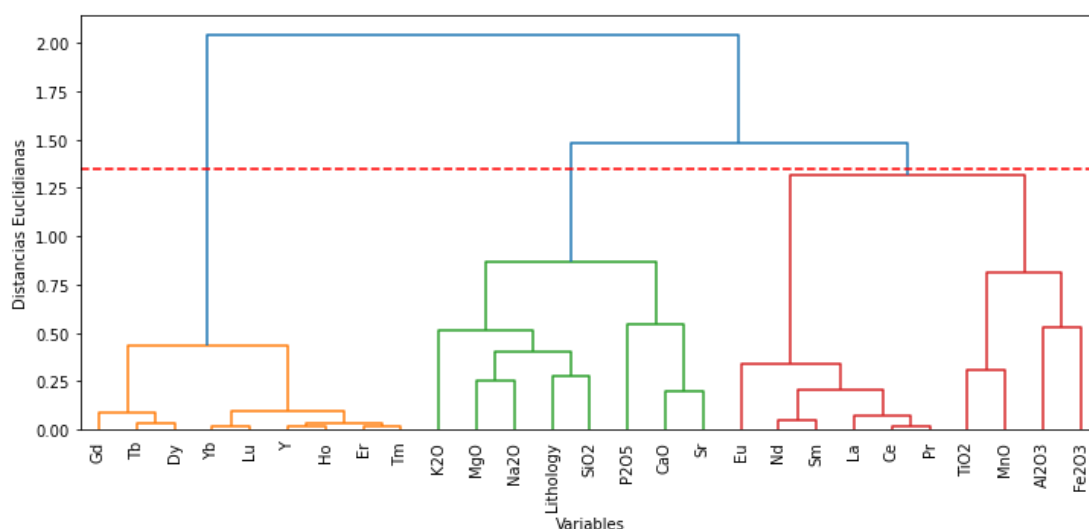


Figura 4.7. Ejemplo de dendrograma jerárquico, el cual muestra la asociación entre 27 elementos de una base de datos. En el eje horizontal se indican los elementos químicos de la base de datos. El eje vertical corresponde a la medida de similitud. El eje horizontal corresponde a las variables agrupadas. La línea roja se usa para marcar el límite de similitud y así determinar el número de clústeres, para este caso un número de 3 clústeres.

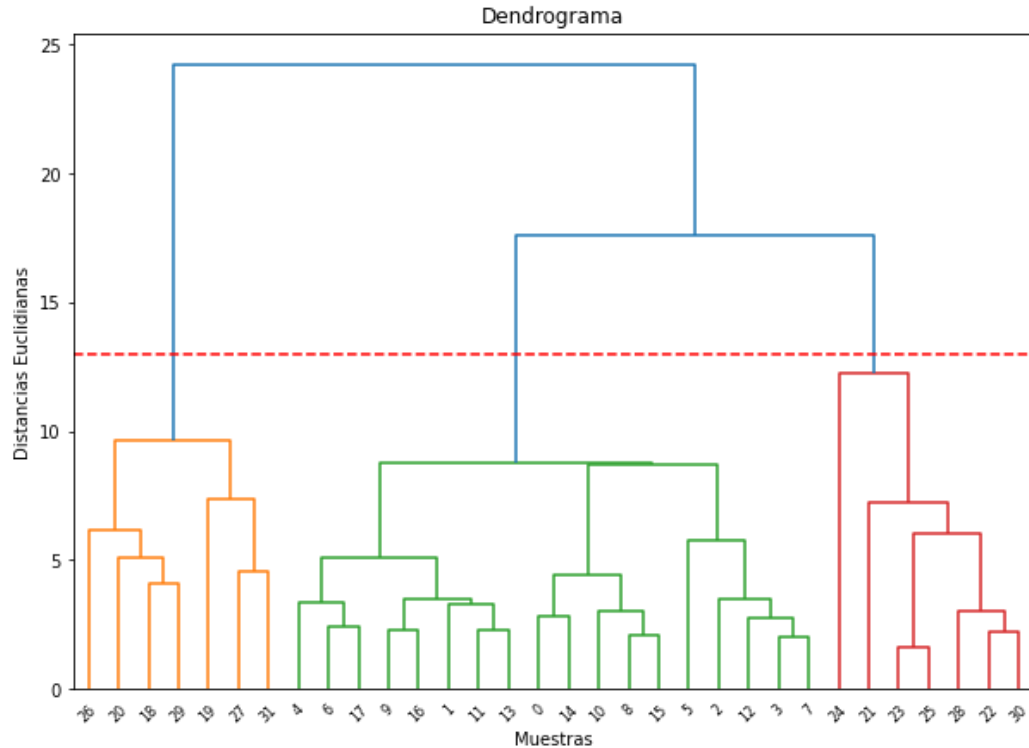


Figura 4.8. Ejemplo de dendrograma jerárquico de clústeres obtenidos, el cual muestra la asociación entre 32 muestras de una base de datos. En el eje horizontal se indican las muestras de la base de datos, enumeradas de 0 a 31. El eje vertical corresponde a la medida de similitud. El eje horizontal corresponde a las muestras. La línea roja se usa para marcar el límite de similitud y así determinar el número de clústeres, para este caso un número de 3 clústeres.

4.7.2 K-means

El algoritmo *K-means* divide un conjunto de datos en clústeres no superpuestos, basados en un conjunto de elementos (MacQueen, 1967; Rahman, 2017). Los clústeres son definidos por un centroide, mientras que la agrupación se consigue minimizando iterativamente el SSE. La determinación del centroide se realiza tomando primero los datos como primer centroide, segundo dato como segundo centroide, y así sucesivamente hasta el número de centroides necesarios. El siguiente paso es calcular la distancia del punto para agrupar a cada centroide disponible y que este agrupado en consecuencia con la distancia más cercana al centroide (Kristanto et al., 2016, citado en Nainggolan et al., 2019; para más detalles ver subsección Tabla 2.1). *K-means* ha demostrado mediante el análisis de muestras de rocas ser capaz de comprender e identificar patrones geoquímicos de la mineralización (Dmitrijeva et al., 2019b; Hajnajafi et al., 2021).

Para determinar el número óptimo de clústeres en el análisis de agrupamiento, se evaluaron cuatro métricas diferentes: el método del "Codo", el coeficiente de *Silhouette*, el índice Davies Bouldin y *Gap Statistics*. La Figura 4.9 ilustra un ejemplo de los gráficos de las métricas empleadas. Los resultados indican que el número óptimo de clústeres se encuentra en el rango de 3 a 5, según el método utilizado. El método del "Codo" sugiere que el número óptimo de clústeres está entre 3 y 5, mientras que el coeficiente de *Silhouette* sugiere que el rango óptimo es entre 2 y 3 (Figura 4.9b). En particular, el valor del coeficiente de *Silhouette* para 2 clústeres es de 0.39, mientras que para 3 clústeres es de 0.33. Por su parte, el índice Davies Bouldin indica que el número óptimo de clústeres se encuentra entre 3 y 4. Finalmente, la estadística Gap muestra que 3 clústeres es una buena elección, ya que las puntuaciones para 4 y 5 clústeres son inferiores. En consecuencia, se decidió optar por 3 clústeres como el número óptimo, debido a que este valor está en el rango sugerido por la mayoría de las métricas evaluadas.

En la tabla 4.2 se presentan los resultados de un ejemplo de la evaluación de la calidad del *clustering*, utilizando los clústeres obtenidos por *K-means* y el Agrupamiento Jerárquico como etiquetas con respecto al elemento La (variable de entrada), para un total de 3 clústeres. Tanto el método de *K-means* como el Agrupamiento Jerárquico tienen un índice ARI de 0.104, lo que indica una similitud muy baja entre los resultados de agrupamiento y las etiquetas verdaderas. El índice AMI para *K-means* es de 0.0736 y para Agrupamiento Jerárquico es de 0.098, lo que sugiere una muy baja similitud entre los resultados de agrupamiento y las etiquetas verdaderas. La *Homogeneity* y *Completeness* para *K-means* son de 0.137 y 0.165, respectivamente, lo que indica que los clústeres no son bastante heterogéneos y no todos los elementos de los clústeres se agruparon correctamente. La *V-measure* para *K-means* es de 0.15, lo que sugiere un rendimiento bajo en términos de precisión y recuperación. Por otro lado, el Agrupamiento Jerárquico presenta una *V-measure* de 0.16, al igual que *K-means* muestra que la calidad de los clústeres formadas es baja. Se puede observar que el Agrupamiento Jerárquico tiene una mayor *Homogeneity* y *Completeness* en comparación con *K-means*, lo que sugiere que podría ser más adecuado para este conjunto de datos.

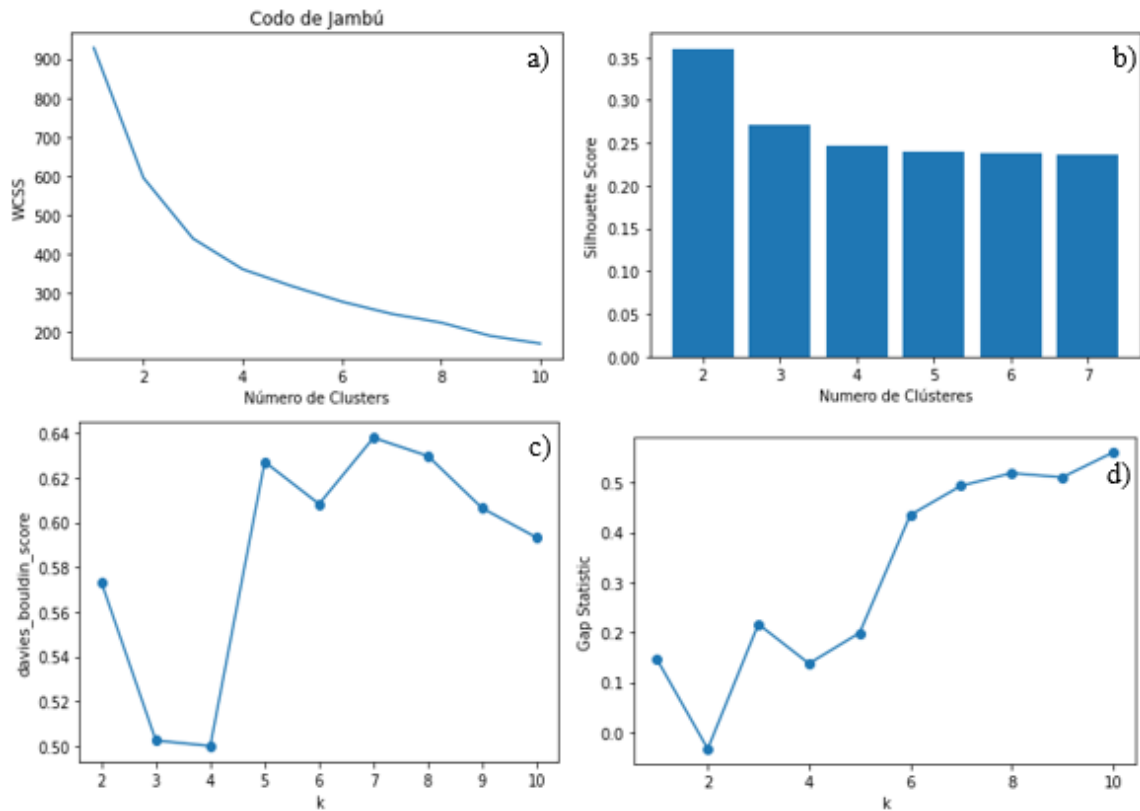


Figura 4.9. Ejemplos de métricas para determinar el número óptimo de clústeres en C-M. a) Método del Codo. b) Método *Silhouette*. c) Método *Davies Boudin Score*. d) Método *Gap Statistic*. El eje Y presenta las puntuaciones para cada método.

Tabla 4.2. Ejemplos de métricas de rendimiento de las propiedades deseables.

	ARI	AMI	Homogeneity	Completeness	V-measure
K-means	0.104	0.0736	0.137	0.165	0.15
AJ	0.101	0.098	0.149	0.194	0.16

Abreviación: Agrupamiento Jerárquico (AJ).

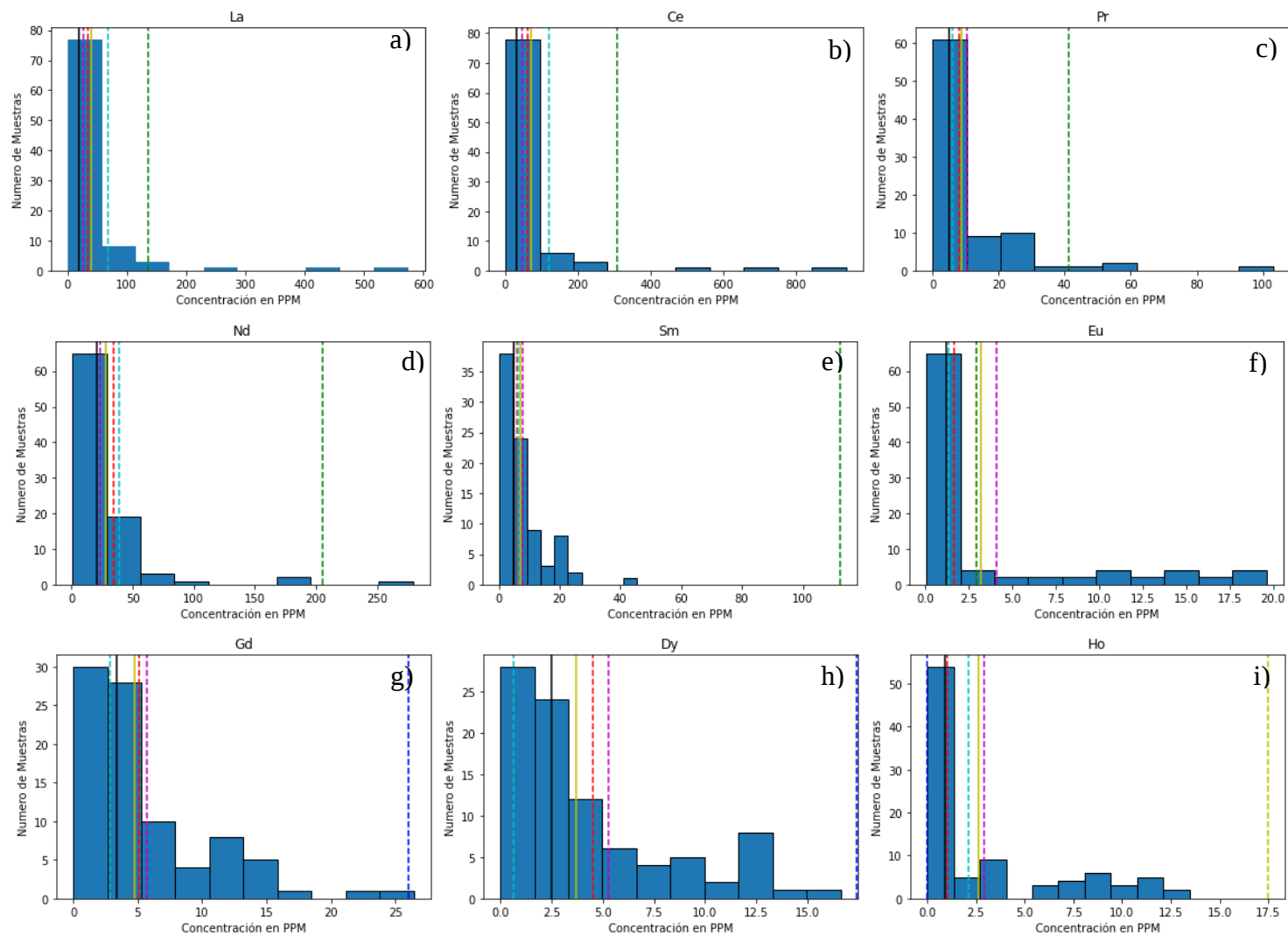
CAPÍTULO 5: RESULTADOS

5.1 Estadística básica de concentraciones geoquímicas

La Tabla 5.1 resume la estadística básica de todos los elementos analizados, mientras la Figura 5.1 muestra los histogramas de frecuencia solo para las REE. La mayoría de las REE en el área de estudio presentan un valor promedio superior a las concentraciones promedio de la corteza (ver Tabla 2.1), a excepción de Dy (en Mantoverde), Ho (en Candelaria-Punta del Cobre), Er (en Mantoverde) y Yb (en Mantoverde). Sin embargo, los histogramas muestran que el grueso de la población de muestras presenta valores de REE que son inferiores o cercanos al valor promedio de la corteza, es decir, se ubican a la izquierda de la línea roja en la Figura 5.1. Los histogramas muestran también que los valores máximos para las REE están en el orden o superan la concentración media de algunos depósitos de referencia, con excepción de los elementos Dy, Ho y Sm donde todos los valores medidos son menores a leyes explotadas (comparar valores máximos con líneas punteadas en Figura 5.1). Sin embargo, cabe notar que las muestras que superan las leyes de depósitos suelen ser una cantidad <10 muestras.

Tabla 5.1. Estadística básica de la base de datos de elementos geoquímicos. Las columnas “25%”, “50%” y “75%” indican los cuantiles en los cuales fueron particionados los datos.

Elemento	Cantidad de muestras	Promedio	Mediana	std	mín	25%	50%	75%	máx
SiO ₂ (%)	81	56.448765	56.37	10.337608	28.66	50.39	56.37	61.65	85.79
TiO ₂ (%)	81	3.154568	0.73	5.753566	0.06	0.52	0.73	1.02	20.17
Al ₂ O ₃ (%)	81	13.614691	14.82	4.249380	3.66	11.38	14.82	17.00	19.70
Fe ₂ O ₃ (%)	81	9.365556	7.00	9.349515	0.47	4.67	7.00	11.18	51.44
MnO (%)	81	0.659012	0.11	1.528287	0.00	0.08	0.11	0.31	8.62
MgO (%)	81	3.186049	2.87	2.068497	0.27	1.67	2.87	4.70	8.34
CaO (%)	81	3.957284	2.92	3.354161	0.04	1.13	2.92	6.43	14.49
Na ₂ O (%)	81	2.301852	1.00	2.265648	0.00	0.20	1.00	3.98	8.43
K ₂ O (%)	81	3.576914	2.77	3.340360	0.01	0.98	2.77	5.30	13.42
P ₂ O ₅ (%)	81	0.209259	0.15	0.265767	0.01	0.10	0.15	0.24	2.20
La (ppm)	81	40.778025	19.90	84.118460	0.60	9.20	19.90	29.90	574.00
Ce (ppm)	81	71.271235	30.50	140.924641	1.40	13.10	30.50	58.60	937.00
Pr (ppm)	81	8.917407	4.99	13.663053	0.00	1.40	4.99	10.60	103.00
Nd (ppm)	81	28.372840	20.32	40.611279	0.70	9.30	20.32	30.40	279.00
Sm (ppm)	81	7.232099	4.90	7.450841	0.20	2.45	4.90	8.50	45.30
Eu (ppm)	81	3.177531	1.21	4.951046	0.05	0.57	1.21	1.97	19.61
Gd (ppm)	81	4.762099	3.38	4.694184	0.00	1.60	3.38	5.40	26.40
Dy (ppm)	81	3.726914	2.52	3.739328	0.00	0.60	2.52	5.00	13.99
Ho (ppm)	81	2.630370	0.90	3.564796	0.00	0.51	0.90	2.90	12.95
Er (ppm)	81	2.680370	1.60	3.122183	0.00	0.60	1.60	2.80	12.89
Tm (ppm)	81	2.009259	0.38	3.249215	0.00	0.24	0.38	1.60	11.47
Yb (ppm)	81	2.577901	1.70	2.986738	0.00	0.70	1.70	2.61	12.73
Lu (ppm)	81	1.927284	0.42	3.022637	0.09	0.28	0.42	1.60	13.53
Y (ppm)	81	19.206543	18.63	11.722056	4.00	10.00	18.63	25.00	67.00
Sr (ppm)	81	237.970988	110.00	233.063098	8.00	52.62	110.00	430.00	820.00



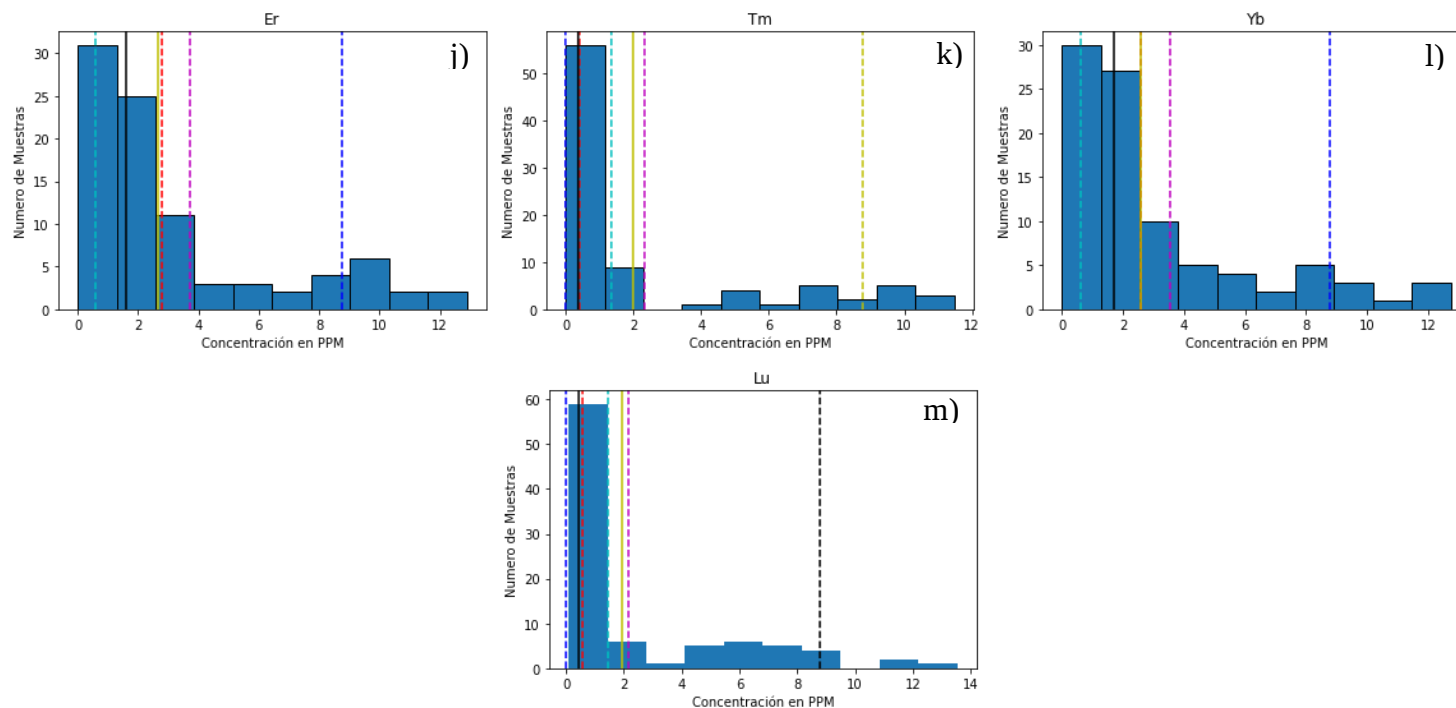


Figura 5.1. Histogramas de frecuencia de REE. La línea segmentada roja indica la concentración promedio del elemento en la corteza (Tabla 2.1). Las líneas de colores representan las concentraciones promedio en diferentes sectores: línea continua amarilla = base de datos de este trabajo (Tabla 5.1), línea continua negra = mediana en la base de datos de este trabajo (Tabla 5.1), línea segmentada magenta = distrito Candelaria-Punta del Cobre (Del Real et al., 2018; Marschik y Fonboté 2001a, b; Marschik et al 1996; 2003), línea segmentada celeste = distrito Mantoverde (Rieger et al., 2010a), línea segmentada verde = depósito Browns Range (Tabla 2.1), línea segmentada azul = depósito Ellior Lake Teasdale (Tabla 2.1), línea segmentada amarilla = depósito Songwe Hill (Tabla 2.1), línea segmentada negra = depósito Mountain Pass (Tabla 2.1).

La Figura 5.2 muestra la matriz de correlación obtenida para toda la población de datos. La matriz de correlación compila los coeficientes de correlación estadística obtenidos para las concentraciones de dos pares de elementos. Los valores cercanos a 1 (colores azulados en Figura 5.2) indican una fuerte relación directa entre los elementos, mientras que valores cercanos a -1 (colores rojizos en Figura 5.2) indican una fuerte relación inversa. En este trabajo valores de correlación menor a 0.3 se consideran una correlación débil, valores de correlación entre 0.3 y 0.7 se consideran una correlación moderada, y valores de correlación mayor a 0.7 se consideran una correlación fuerte. Estos valores son ampliamente aceptados en la literatura científica como criterios para evaluar la fuerza de la correlación entre variables (Cohen, 1988; Pallant, 2020; Mukaka, 2012).

De la matriz de correlación, se destacan los siguientes resultados:

- Tanto las HREE como las LREE presentan correlación fuerte (>0.7) solo con elementos de su mismo sub-grupo, con excepción del Pr (una LREE que solo presenta una correlación >0.7 con el Gd), y el Eu (una LREE que presenta también correlaciones >0.7 con algunas HREE).
- En el caso de los óxidos mayores, el Na_2O presenta una correlación positiva moderada (0.4-0.6) con los HREE, Sm, Eu, Y, Sr, Al_2O_3 y CaO . El Al_2O_3 presenta una correlación moderada con Na_2O , Eu y Sr. En el caso del Fe_2O_3 , solo destaca su fuerte correlación negativa con el SiO_2 . El SiO_2 tiene una correlación negativa moderada con las LREE. El TiO_2 presenta una correlación negativa moderada con K_2O y una correlación fuerte con MnO . El K_2O presenta en general bajas correlaciones con todos los elementos, a excepción del Al_2O_3 .

5.2 Análisis de Componentes Principales (PCA)

La Figura 5.3 muestra gráficamente la influencia o peso de cada variable (elementos químicos) en cada una de las componentes principales obtenidas mediante el análisis PCA. En la matriz de la Figura 5.3, cada fila representa una variable, mientras las columnas corresponden a cada componente principal. El grado de influencia de una variable en un componente principal se muestra mediante un color de la celda donde coinciden la correspondiente fila y columna.

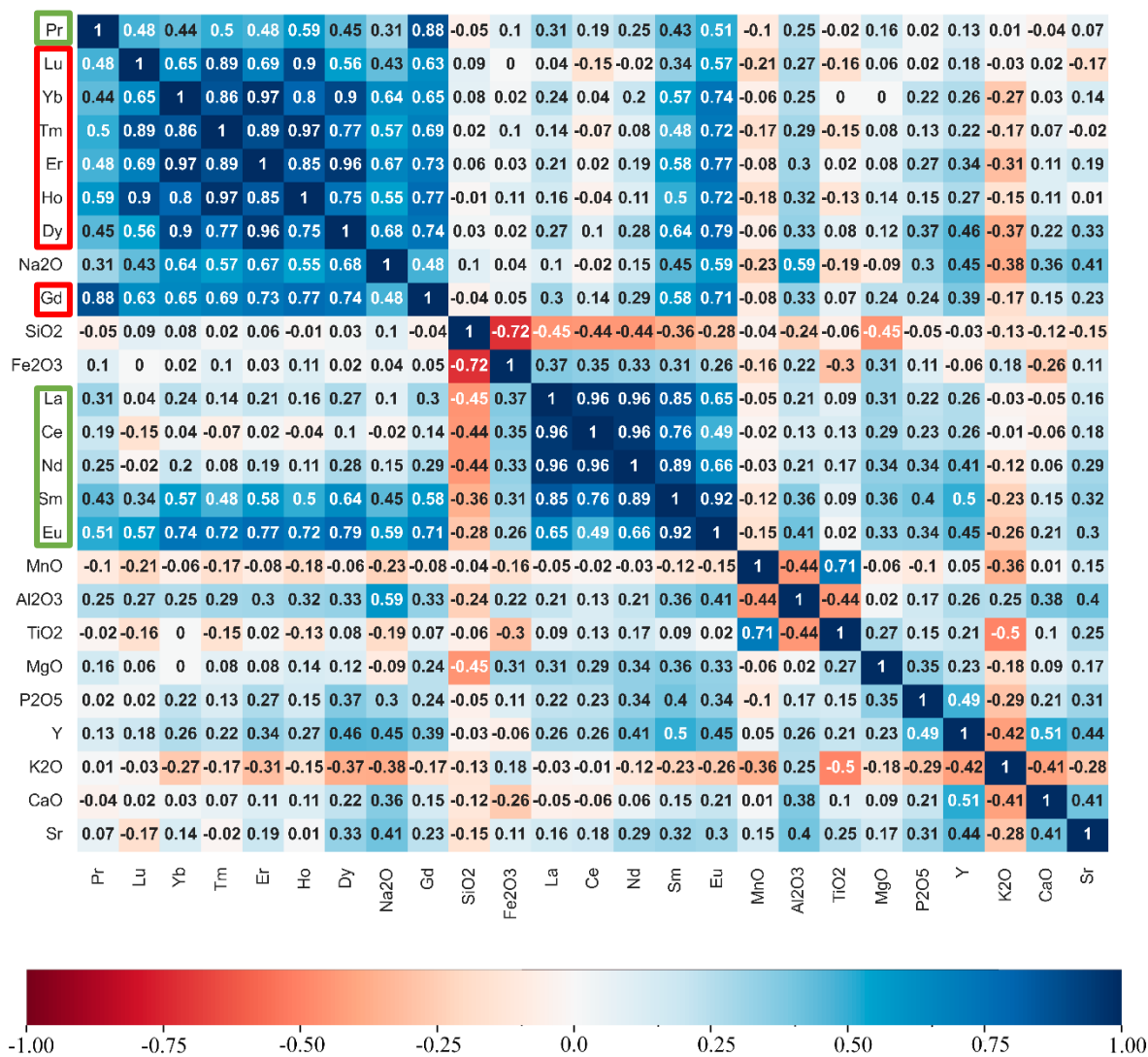


Figura 5.2. Matriz de correlación ordenada. Los rectángulos verdes indican las LREE, mientras que los rectángulos rojos indican las HREE.

El valor del grado de influencia de una variable en una componente principal varía en un rango de -1 a 1, donde valores cercanos a 1 ó -1 indican una fuerte influencia, mientras que valores cercanos a 0 indican una baja influencia. La Figura 5.4 muestra el porcentaje de la varianza acumulada explicada por la adición de cada componente principal. La primera componente explica el 34% y el segundo componente explica el 17% de la varianza observada. Si se evalúan únicamente las dos primeras componentes se conseguiría explicar el ~52% de la varianza observada. Debido a que se trata de simplificar la interpretación de los datos y reducir la complejidad, se decide trabajar con las dos

primeras componentes, ya que tienen una importante contribución en la explicación de la variabilidad en los datos.

La Figura 5.3 indica que la componente principal que explica la mayor varianza (PC1) está principalmente influenciada por las HREE y Eu, y en menor medida por Na₂O, Pr y Sm. Por su parte la componente principal PC2 está principalmente influenciado por las restantes LREE, y levemente por Fe₂O₃. La componente principal PC3 está principalmente influenciada por MnO y TiO. El PC4 está influenciado principalmente por CaO, Al₂O₃ y Sr. El PC5 está influenciado principalmente por Fe₂O₃, MgO y SiO₂.

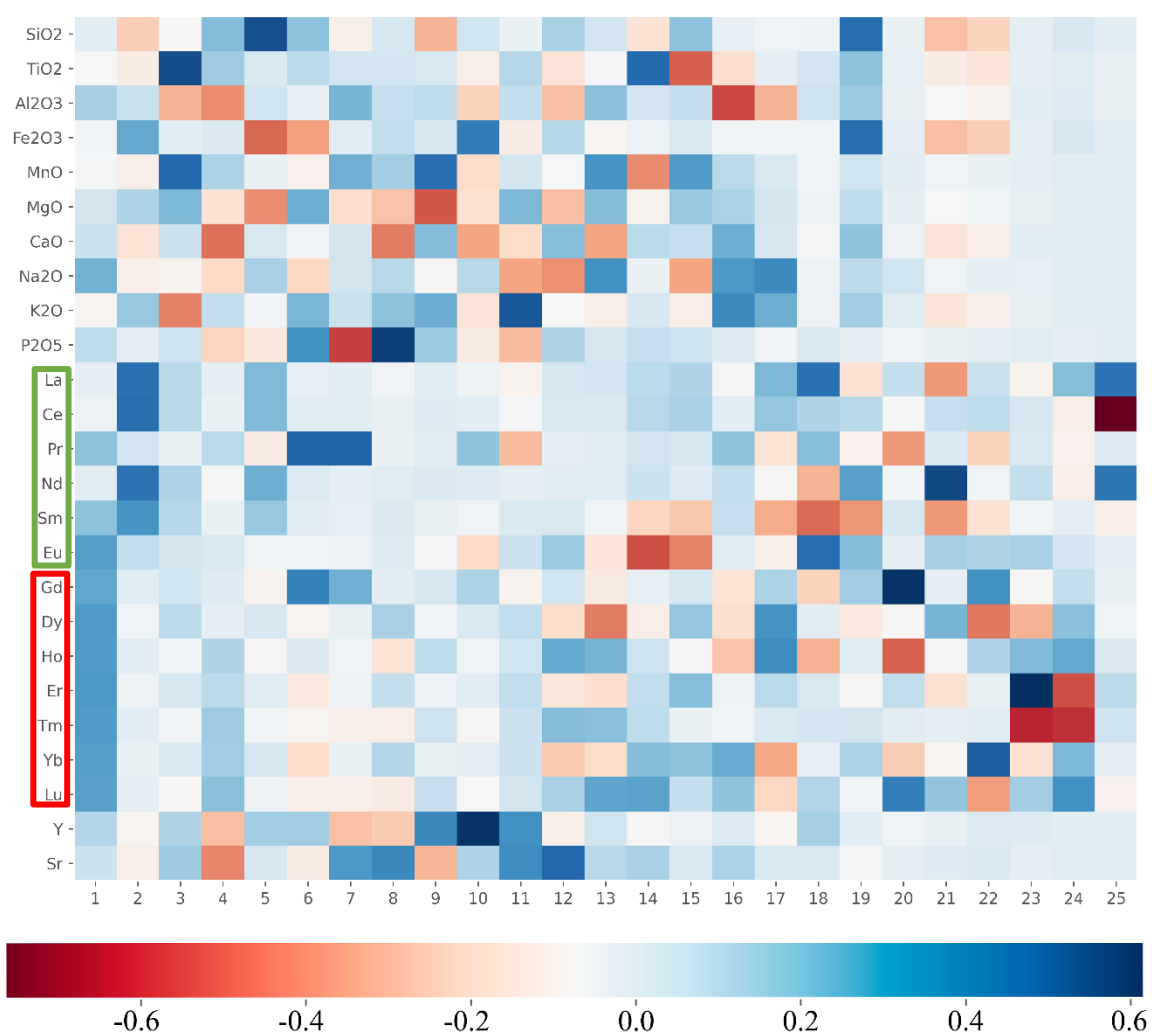


Figura 5.3. Mapa de calor que muestra la influencia de las variables en cada componente principal obtenida. El eje vertical corresponde a las variables (elementos químicos). El eje horizontal corresponde a los componentes principales. Los rectángulos verdes indican las LREE, mientras que los rectángulos rojos indican las HREE.

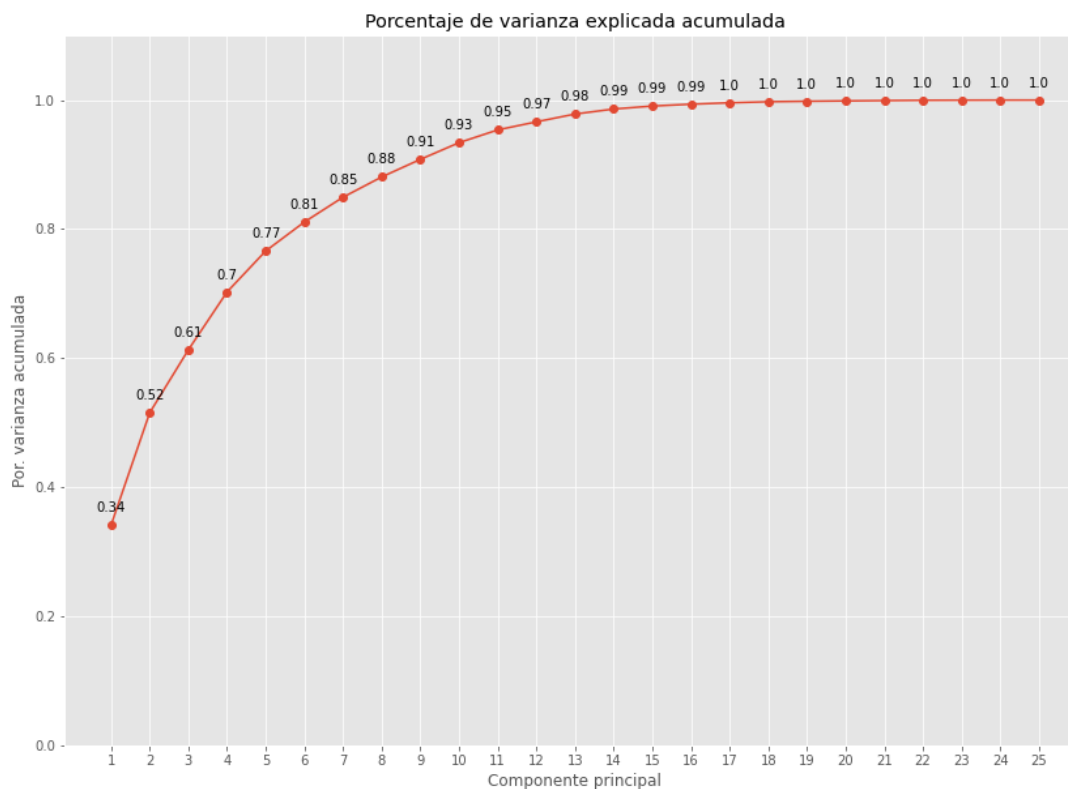


Figura 5.4. Curva que muestra el porcentaje de varianza acumulada explicada al considerar cada componente principal.

5.3 Agrupamiento Jerárquico

Para identificar las relaciones entre las variables se construyó un dendrograma de similitud mediante el Agrupamiento Jerárquico. La Figura 5.5 presenta el dendrograma de clústeres de las variables derivada de la base de datos del área de estudio. La línea roja segmentada horizontal indica la cantidad de grupos o clústeres que se considerará para el análisis de los datos, que en este caso corresponden a tres clústeres (C0, C1 y C2 en la Figura 5.5), ya que se logra una separación más uniforme de las variables (elementos químicos). El clúster C0 comprende la mayor parte de los óxidos mayores además de Sr e Y. El clúster C1 agrupa a todas las HREE en conjunto con Pr y NaO₂. Por su parte el clúster C2 agrupa tanto las LREE como SiO₂ y Fe₂O₃.

También se realizó un agrupamiento jerárquico de las muestras del área de estudio (Figura 5.6). De igual manera que el agrupamiento jerárquico de las variables, se trazó una línea roja para delimitar la cantidad de clústeres que serán evaluados, que para el caso de la Figura 5.6 se seleccionaron cinco clústeres. Además, para determinar el número óptimo

de clústeres se utilizaron las métricas Calinski-Harabasz, Cophenética y el índice Davies-Bouldin, y se obtuvo que cinco clústeres son los óptimos (Tabla 5.1). Estas métricas son discutidas con mayor profundidad en la sección 6.2. Los clústeres 0 y 1 agrupan muestras pertenecientes a ambos distritos mineros estudiados, mientras que los restantes clústeres solo presentan muestras del distrito Candelaria-Punta del Cobre.

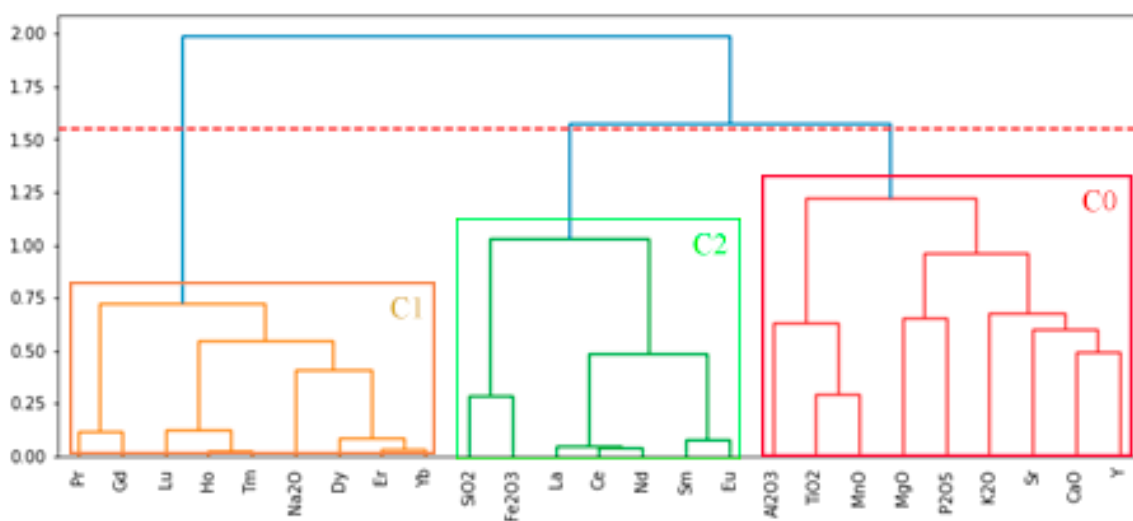


Figura 5.5. Dendrograma de Agrupación Jerárquica para las variables (elementos químicos).

Las Figuras 5.7 y 5.8 muestran la distribución espacial de los clústeres de muestras en el distrito Candelaria-Punta del Cobre y Mantoverde respectivamente. El clúster C0 es el que contiene mayor cantidad de muestras (26) y está presente tanto en el distrito Candelaria-Punta del Cobre como en el distrito Mantoverde, aunque con mayor presencia en este último (21 muestras). De igual manera el clúster C1 (11 muestras) tiene presencia en ambos distritos, con seis muestras en Candelaria-Punta del Cobre y cinco en Mantoverde. Los clústeres C2 (13 muestras), C3 (13 muestras) y C4 (18 muestras) solo tienen presencia en el distrito Candelaria-Punta del Cobre.

Tabla 5.2. Métricas número de clústeres para Agrupamiento Jerárquico

Nº Clústeres	CH	DB	Cophenética
5	51.138	0.948	0.639
2	44.57	0.922	0.639

Abreviaciones: Calinski-Harabasz (CH), Davies-Bouldin (DB).

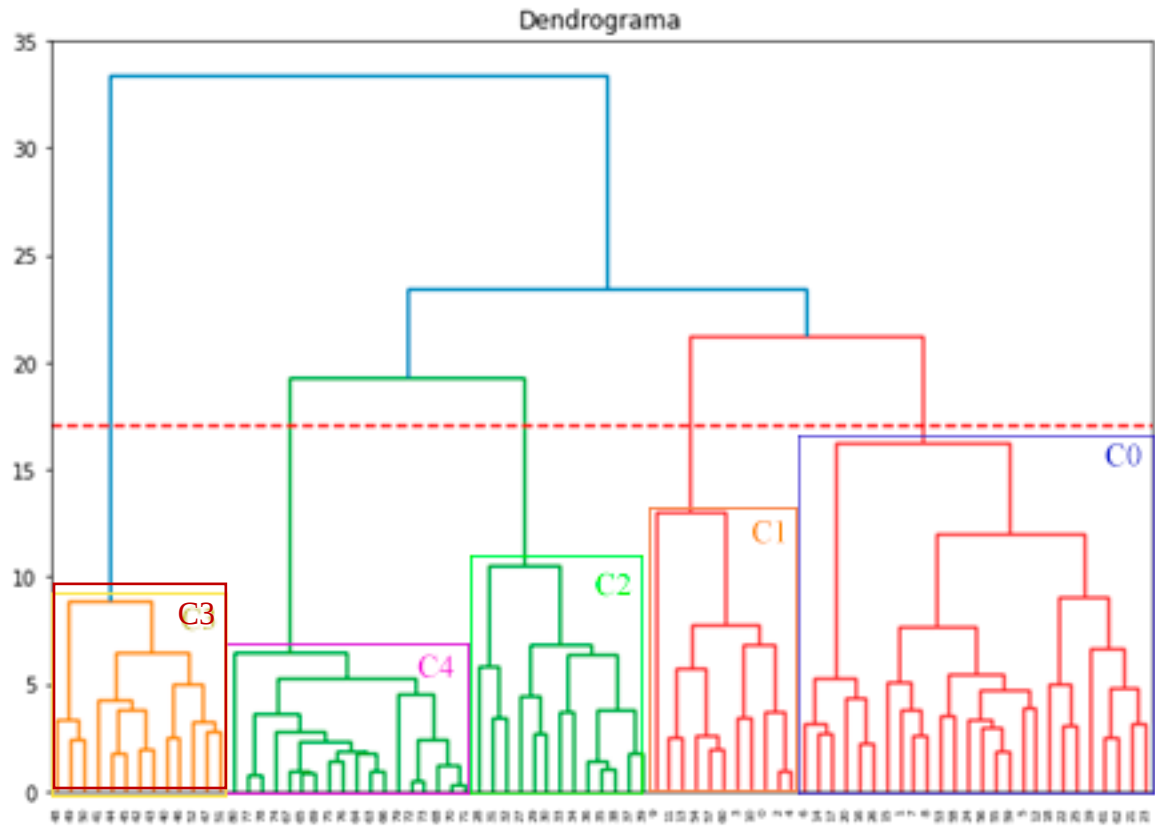


Figura 5.6. Dendrograma de las muestras. La línea roja segmentada marca el límite de clústeres seleccionados (cinco). Los clústeres están indicados mediante cuadrados y rectángulos con sus respectivos colores.

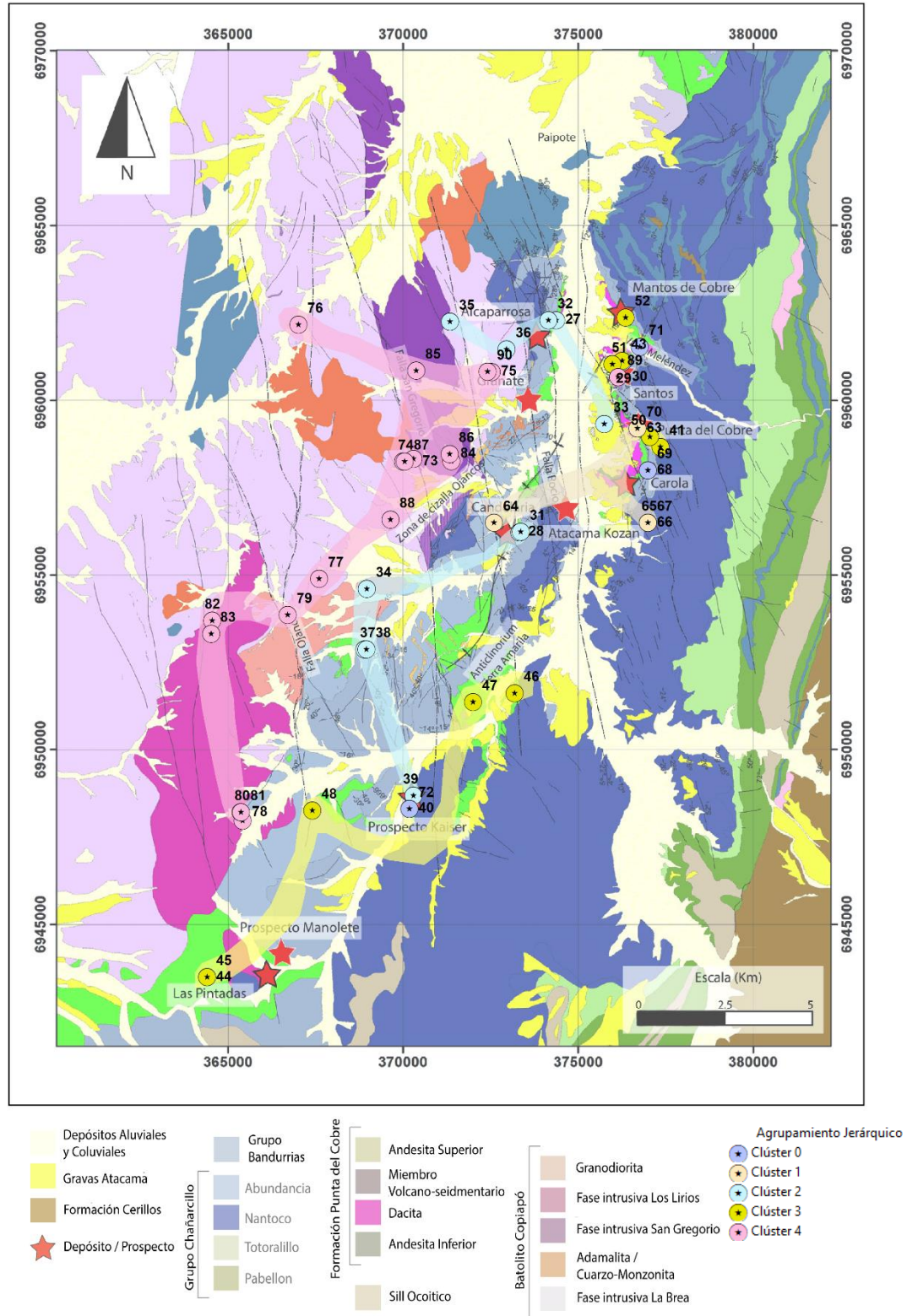


Figura 5.7. Distribución espacial de los clústeres obtenidos mediante el Agrupamiento Jerárquico de las muestras en el Distrito Candelaria-Punta del Cobre (mapa geológico tomado de Del Real et al., 2018).

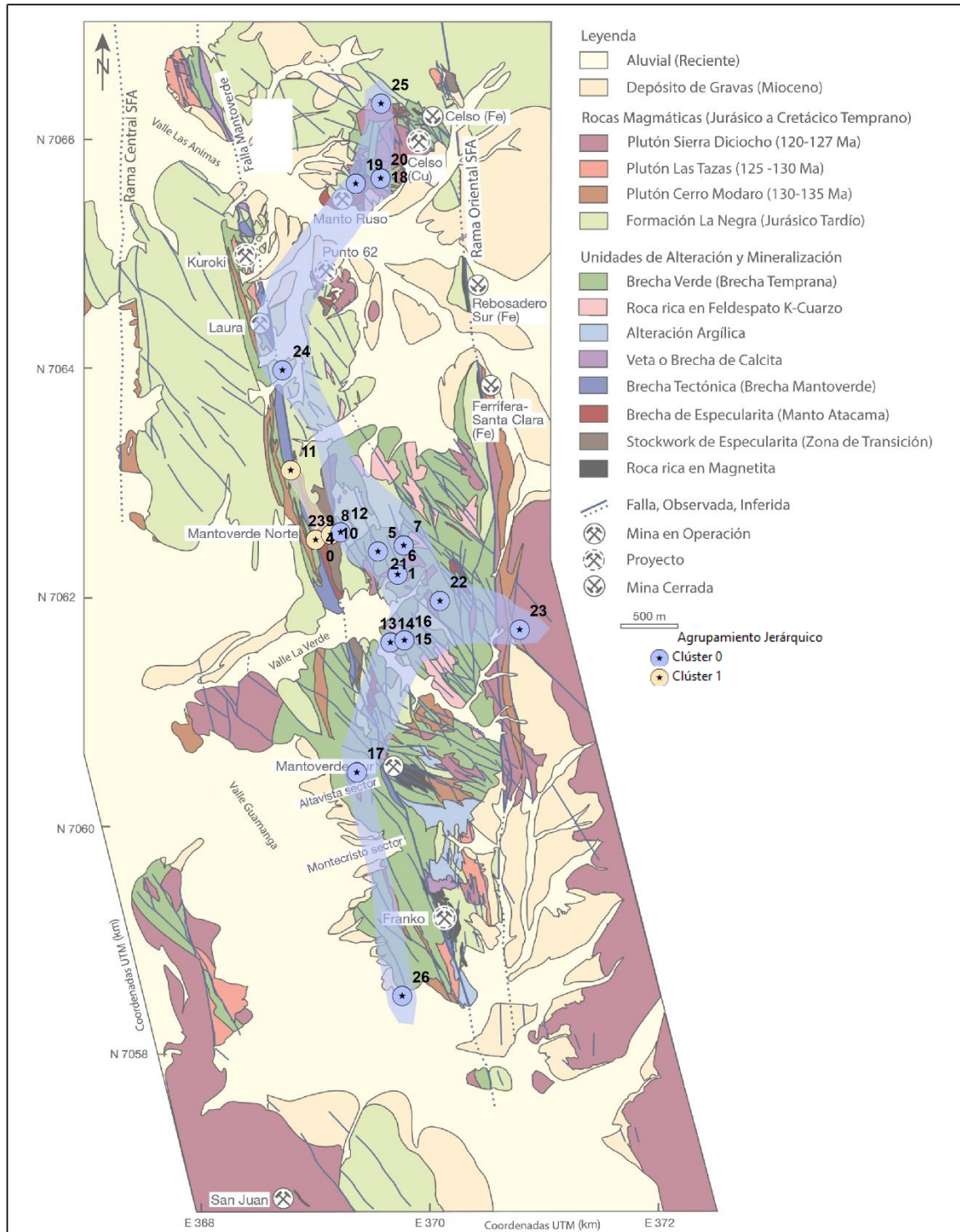


Figura 5.8. Distribución espacial de los clústeres obtenidos mediante el Agrupamiento Jerárquico en el Distrito Mantoverde (mapa geológico tomado de Rieger et al., 2010a).

5.4 *K-means*

El óptimo de las agrupaciones de *K-means* se determinó mediante las métricas del Codo, el índice Davies Bouldin, *Gap Statistic* y *Silhouette*. El gráfico del Codo, el índice Davies Bouldin y *Gap Statistic* (Figura 5.9) se consiguieron ejecutando *K-means* en el conjunto de datos para un rango de valores de dos a diez, por su parte *Silhouette* se ejecutó en un rango de valores de dos a siete. Las métricas en su mayoría indican que el número de clústeres óptimos está entre dos y cinco. El método del Codo muestra que de 4 a 6 clústeres pueden ser óptimos. *Silhouette* indica que dos y cinco clústeres son óptimos, y el índice Davies Bouldin indica que el óptimo es de cinco clústeres. *Gap Statistic* muestra que sobre cinco clústeres la disminución de la variabilidad por clúster no aumenta de manera significativa.

En virtud de lo anterior, se analizaron los casos de cinco clústeres. En la sección 6.2 se discute con mayor profundidad la cantidad de clústeres escogidos. Las Figuras 5.10 y 5.11 muestran la distribución espacial en los distritos Candelaria-Punta del Cobre y Mantoverde de los cinco clústeres obtenidos con *K-means*. En el distrito Candelaria-Punta del Cobre el clúster 0 es el de mayor presencia, se concentra principalmente en la zona de Dacitas, Fases intrusivas Los Lirios y San Gregorio. El clúster 1 se concentra mayoritariamente en el Grupo Bandurrias y la Formación Abundancia. El clúster 2 presenta escasa presencia y se concentra en la Formación Nantoco. El clúster 3 tiene presencia en el Grupo Bandurrias y Fase Intrusiva San Gregorio. El Clúster 4 tiene presencia en Formación Nantoco y Grupo Bandurrias. Por su parte el distrito Mantoverde solo tiene la presencia de tres clústeres. El clúster 0 está distribuido en el Plutón Sierra Dieciocho. El clúster 2 es el de mayor presencia y se encuentra distribuido en Brecha verde y Formación La Negra. Por último, el clúster 4 tiene presencia en Brecha de Especularita y Brecha Tectónica.

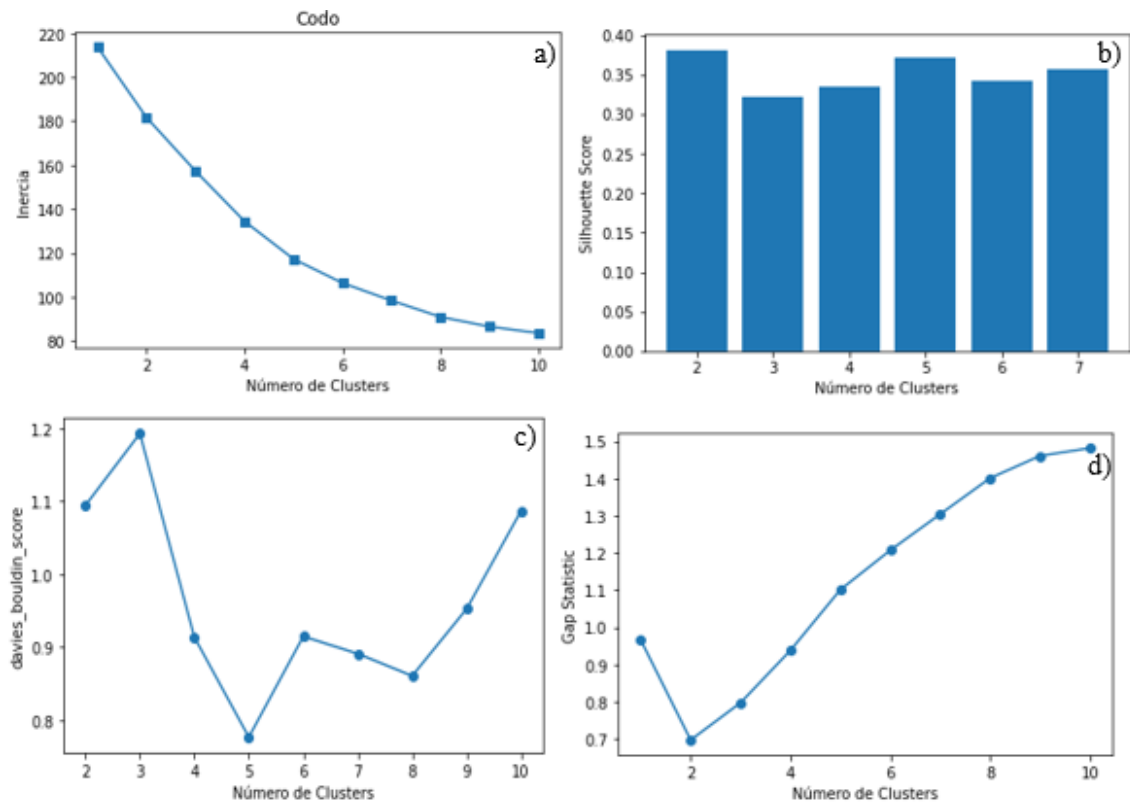


Figura 5.9. Métricas para determinar el número óptimo de clústeres. a) Método del Codo. b) Método *Silhouette*. c) Índice Davies Bouldin. d) Método *Gap Statistic*. El eje Y presenta las puntuaciones para cada método.

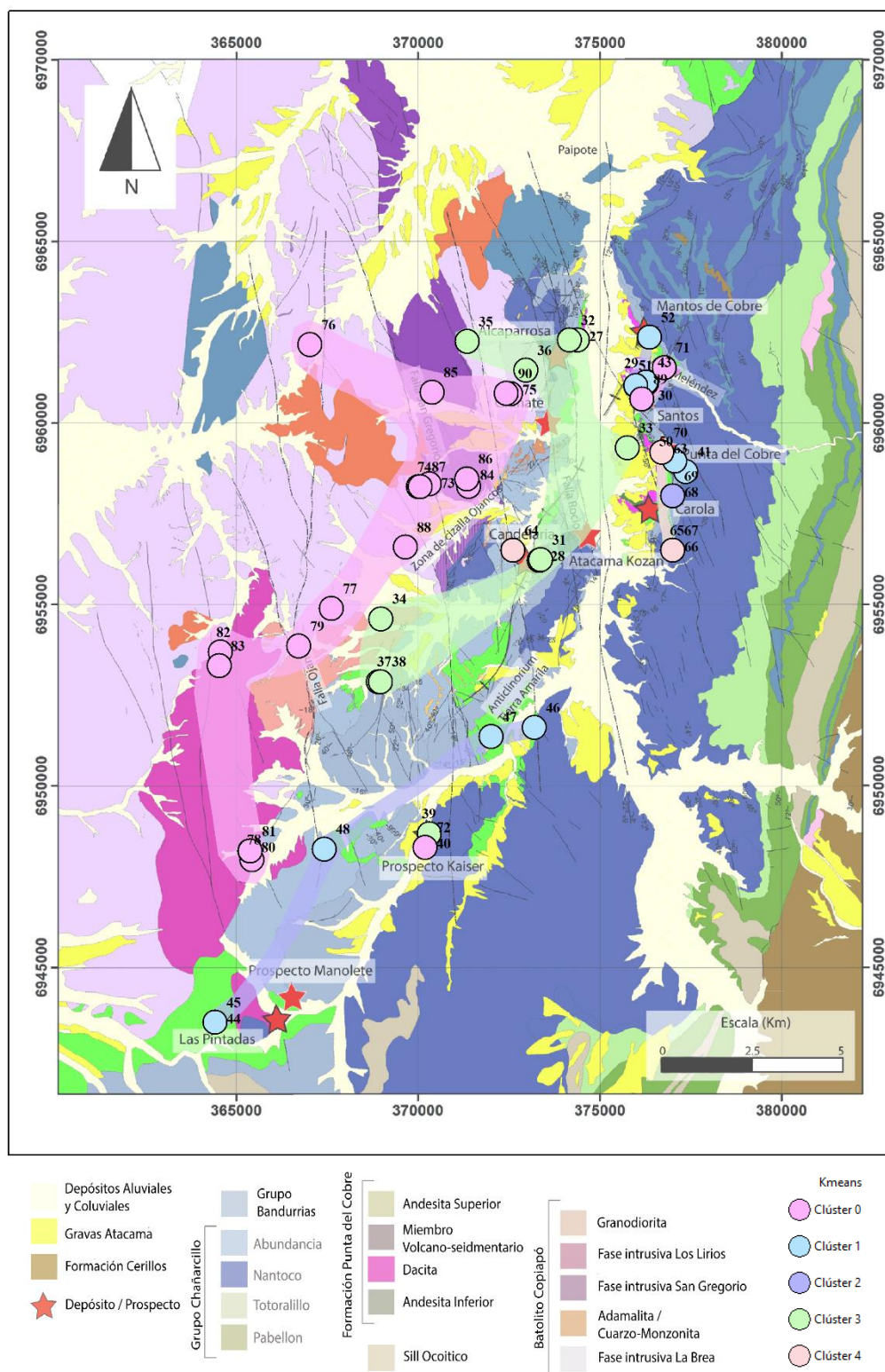


Figura 5.10. Distribución espacial de los clústeres obtenidos mediante *K-means* (asumiendo agrupamiento en cinco clústeres) en el Distrito Candelaria-Punta del Cobre (geología tomada de Del Real et al., 2018).

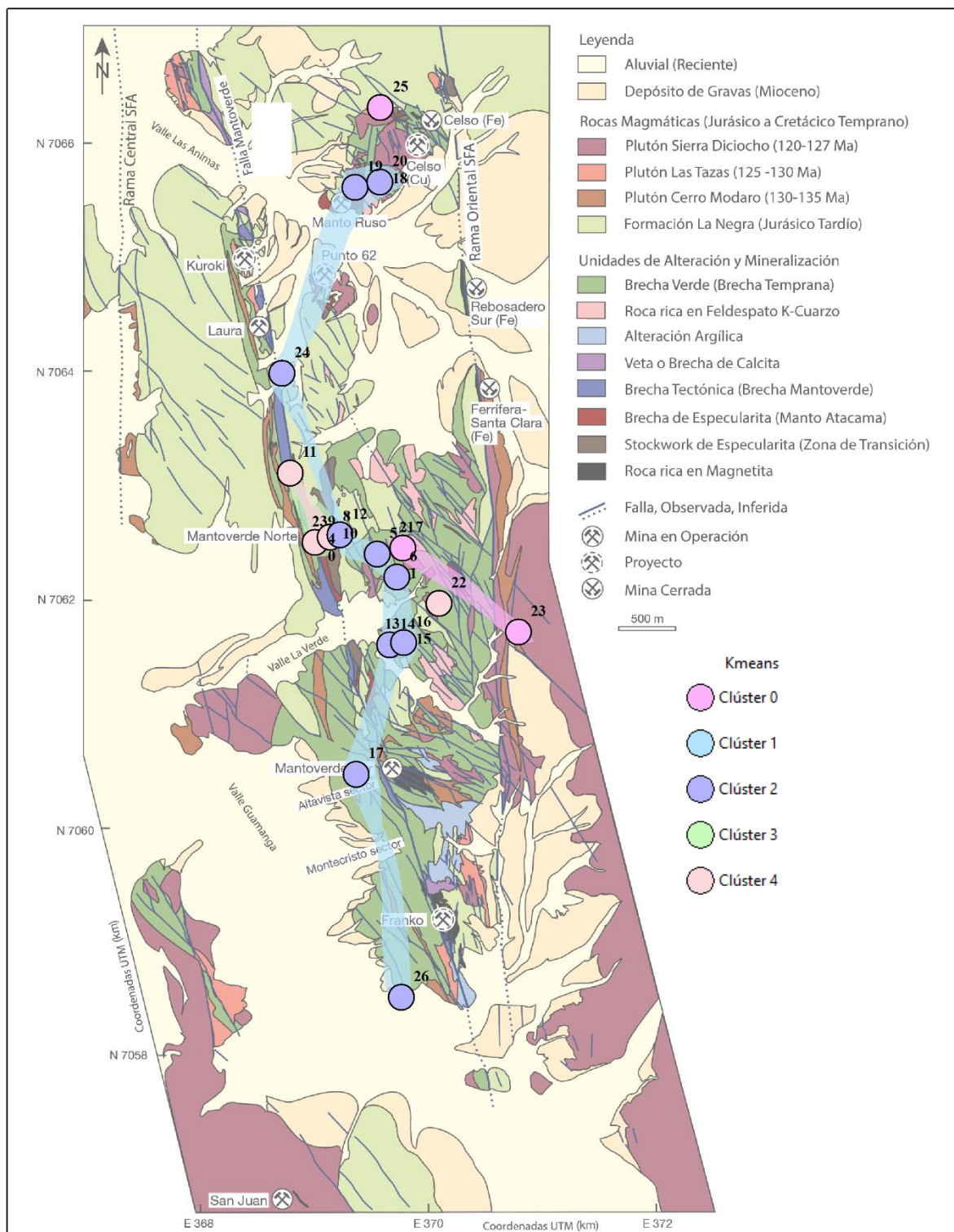


Figura 5.11. Distribución espacial de los clústeres obtenidos mediante *K-means* (asumiendo agrupamiento en cinco clústers) en el Distrito Mantoverde (modificado de Rieger et al., 2010a).

CAPÍTULO 6: DISCUSIÓN

6.1 Elección de cantidad de clústeres para Agrupamiento Jerárquico y *K-means*

Se utilizaron diversas métricas para determinar la cantidad óptima de clústeres en los algoritmos Agrupamiento Jerárquico y *K-means* (ver descripción de estas métricas en sección 2.4.6). Para evaluar los resultados de Agrupamiento Jerárquico se utilizaron las métricas de Calinski-Harabasz, Cophenética y Davies Bouldin. Para evaluar los resultados de *K-means* se emplearon las métricas del Codo, *Silhouette*, Distorsión y Davies Bouldin. La tabla 6.1 muestra los valores obtenidos con cada métrica para una cantidad de 5 y 2 clústeres.

En el caso de Agrupamiento Jerárquico, con cinco clústeres se obtuvo mejores puntuaciones en las métricas Calinski-Harabasz y en el índice Davies Bouldin. Sin embargo, la métrica de Cophenética muestra igual valor para dos y cinco clústeres, lo que sugiere que ambos agrupamientos tienen una capacidad similar de preservar la relación original de las distancias entre las observaciones. Para *K-means*, tanto las métricas de Distorsión como de Davies Bouldin obtienen valores menores con 5 clústeres, lo cual sugiere que esta cantidad de clústeres es más coherente. En contraposición, una agrupación con dos clústeres presenta un mejor valor de *Silhouette*, aunque la diferencia no es significativa.

Después de evaluar los resultados obtenidos por las métricas utilizadas, se decidió utilizar cinco clústeres para los algoritmos *K-means* y Agrupamiento Jerárquico. Esta decisión se basó también en la variabilidad de los datos geoquímicos y en la interpretación de los patrones y tendencias presentes en los datos. Considerando que las concentraciones y asociaciones de elementos químicos en rocas pueden deberse a la sumatoria de varios procesos (mineralogía primaria, eventos hidrotermales, meteorización, etc.), se considera que cinco clústeres podrían eventualmente reflejar esa diversidad de procesos más que solo dos clústeres. También un número de clústeres de cinco resulta más útil dado que permite identificar conjunto de elementos con un sentido geológico más coherente. Además, es el número de clústeres obtenidos por el Agrupamiento Jerárquico lo que

permitirá una comparación más precisa entre los resultados obtenidos por los dos algoritmos.

Tabla 6.1. Métricas de evaluación de *clustering* para Agrupamiento Jerárquico y *K-means* para dos y cinco clústeres.

Algoritmo	Nº Clústeres	Silhouette	CH	DB	Cophenética	Distorsión
AJ	5	-	51.138	0.948	0.639	-
AJ	2	-	44.57	0.922	0.639	-
<i>K-means</i>	5	0.37	-	0.77	-	9019
<i>K-means</i>	2	0.38	-	1.09	-	22446

AJ = Agrupamiento Jerárquico, CH = Calinski-Harabasz, DB = Davies Bouldin.

6.2 Validez de los resultados

Una problemática presente en los algoritmos de aprendizaje no supervisado es la dificultad o imposibilidad de validar la exactitud de los resultados, ya que por la naturaleza de los problemas abordados no se cuenta con un conjunto de datos que representen el valor o respuesta correcta de lo que se está buscando. A pesar esto, varios estudios han propuesto maneras de evaluar la calidad y validez de los resultados utilizando diferentes métricas (Feizollah et al., 2014; Siless et al., 2013; Rosenberg y Hirschberg, 2007; Li y Rowley, 2002). En este caso, se utilizaron las métricas indicadas en la sección 2.4.7, las cuales se implementaron para evaluar la calidad de la agrupación en los algoritmos Agrupamiento Jerárquico y *K-means*. Si bien las métricas de rendimiento son una parte fundamental en la evaluación de modelos y algoritmos de aprendizaje automático, es importante tener en cuenta que en estas métricas no existe un valor umbral fijo que indique que un resultado es "malo", "moderado" o "fuerte" en todos los casos (Sokolova et al., 2006; Kuhn y Johnson, 2013).

Para este trabajo se utilizaron las siguientes categorías para interpretar los valores de las métricas de rendimiento:

- ARI: valores menores a 0.2 similitud muy baja, valores entre 0.2-0.5 similitud moderada, y valores mayores a 0.5 similitud fuerte (Moradi et al., 2012; Müller y Guido, 2016; Ribeiro y Rios, 2021).

- AMI: valores menores a 0.25 similitud muy baja, valores entre 0.25-0.6 similitud moderada, y valores mayores a 0.6 similitud fuerte (Meilă, 2005; Amigó et al., 2009).
- *V-measure*: valores menores a 0.4 baja, valores entre 0.4-0.8 moderada, y valores mayores a 0.8 fuerte (Rosenberg y Hirschberg, 2007; Amigó et al., 2009).
- En el caso de *Homogeneity* y *Completeness* los autores (Rosenberg y Hirschberg, 2007; Amigó et al., 2009) no establecen valores específicos para cada una de ellas de manera individual, sino que combinan ambas medidas en *V-measure*.

Para las métricas de rendimiento se utilizó La para las LREE y Dy para las HREE porque ambos son elementos representativos de sus respectivos grupos. En la Tabla 6.2 se presentan los resultados de las métricas de evaluación para las LREE utilizando los clústeres obtenidos por *K-means* y el Agrupamiento Jerárquico como etiquetas con respecto a una variable de entrada, que en este caso fue el elemento La. Los resultados muestran una correspondencia moderada entre los clústeres formados y la asignación conocida, según la métrica ARI y una similitud moderada entre las etiquetas reales y las generadas por el algoritmo, según AMI. La métrica *Completeness* muestra que ~41% y ~47% de las muestras han sido asignadas correctamente en los clústeres formados por *K-means* y Agrupamiento Jerárquico, respectivamente. *Homogeneity* indica que los clústeres contienen una moderada cantidad de puntos de una misma clase, aproximadamente el 34% y 40% para *K-means* y Agrupamiento Jerárquico, respectivamente. Un rango de ~40-50% en ambas medidas sugiere que los elementos dentro de cada clúster pueden tener diferencias significativas entre sí, lo que indica que los clústeres no están muy bien definidos o que no se han agrupado adecuadamente los elementos similares. Por último, la métrica *V-measure* muestra que la calidad de los clústeres formados es baja, con una entropía entre *Homogeneity* y *Completeness* de aproximadamente 0.37 y 0.43 para *K-means* y Agrupamiento Jerárquico, respectivamente.

Se efectuó el mismo procedimiento para HREE utilizando como variable de entrada el elemento Dy (Tabla 6.3). Los valores obtenidos son similares a los obtenidos en la Tabla 6.2, aunque ligeramente mayores. Comparando los resultados obtenidos por las métricas para ambos casos, se aprecia que tienen un desempeño moderado en la clasificación de la

variable de entrada. Sin embargo, el algoritmo de Agrupamiento Jerárquico tiene un desempeño ligeramente mejor en términos de ARI, AMI, *Homogeneity*, *Completeness* y *V-measure*. Es importante destacar que la calidad del agrupamiento puede variar dependiendo de las métricas utilizadas para evaluar los resultados.

Tabla 6.2. Métricas de rendimiento de las propiedades deseables para *K-means* y Agrupamiento Jerárquico vs La.

	ARI	AMI	Homogeneity	Completeness	V-measure
<i>K-means</i>	0.221	0.324	0.344	0.419	0.378
AJ	0.228	0.388	0.402	0.478	0.437

Tabla 6.3. Métricas de rendimiento de las propiedades deseables para *K-means* y Agrupamiento Jerárquico vs Dy.

	ARI	AMI	Homogeneity	Completeness	V-measure
<i>K-means</i>	0.256	0.364	0.401	0.422	0.411
AJ	0.347	0.43	0.464	0.483	0.474

6.3 Análisis de resultados

Para referirse a los clústeres obtenidos por cada método, se asignaron los siguientes códigos:

- Agrupamiento Jerárquico: AJC0, AJC1, AJC2, AJC3, AJC4.
- K-means: KMC0, KMC1, KMC2, KMC3, KMC4.

Con el objetivo de visualizar la distribución y enriquecimiento de REE en los clústeres, e identificar la influencia de los óxidos mayores, se graficaron *boxplots* de los clústeres de cada elemento para cada algoritmo. En base a los resultados de la matriz de correlación y a lo observado en los *boxplots*, se escogió graficar los elementos que ilustran de mejor manera el comportamiento de las REE, así como también los óxidos que presentan una mayor correlación con las REE. Siguiendo estos criterios, se escogió La como elemento representativo de las LREE, Eu debido a su correlación tanto LREE como con HREE, Dy para las HREE, y el Y debido a que suele ser agrupado junto con las REE. Las Figuras 6.1 y 6.2 muestran los *boxplots* de los elementos escogidos para el Agrupamiento Jerárquico y *K-means*, respectivamente. Los *boxplots* de los elementos restantes se encuentran en los

Anexos 3 y 4. Para el posterior análisis se utilizó el promedio en ppm de las LREE y HREE. Las Tablas 6.4 y 6.5 resumen las características obtenidas a partir de Agrupamiento Jerárquico y *K-means*, respectivamente.

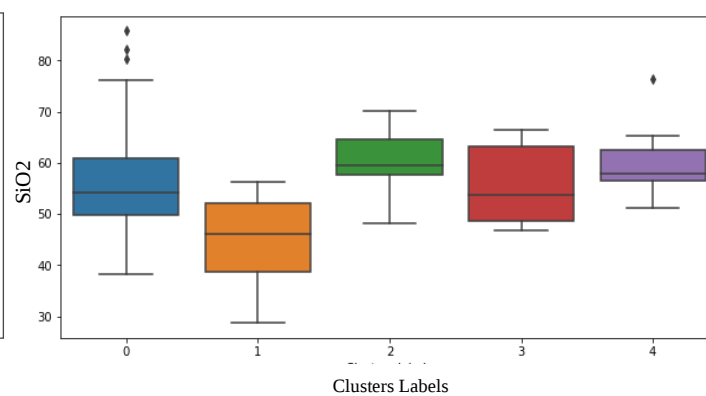
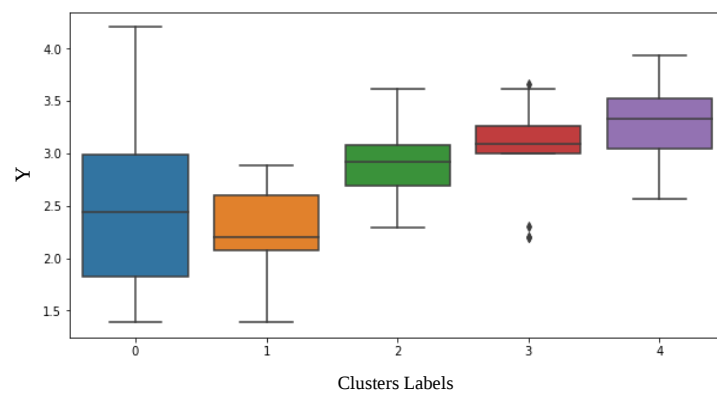
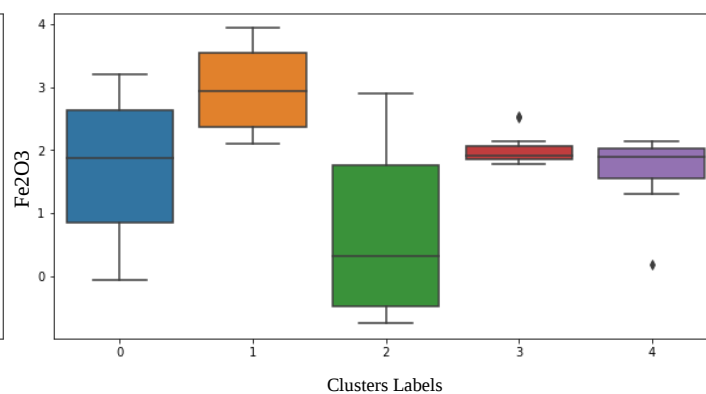
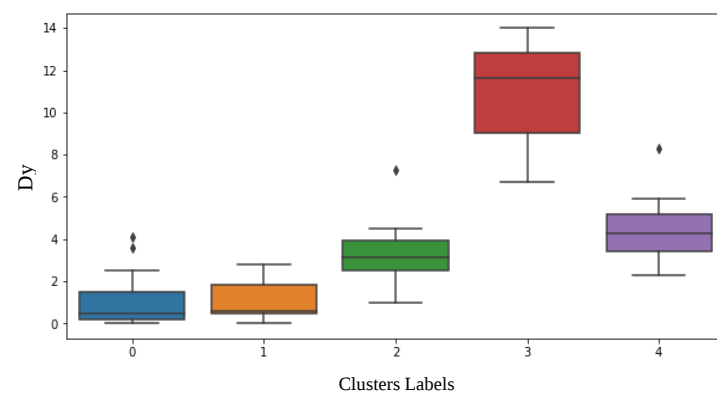
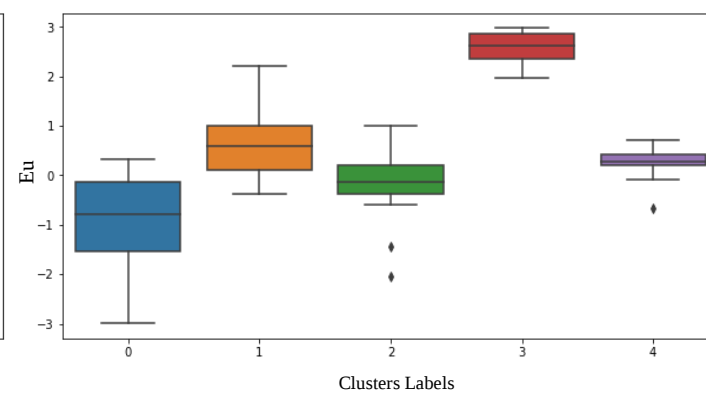
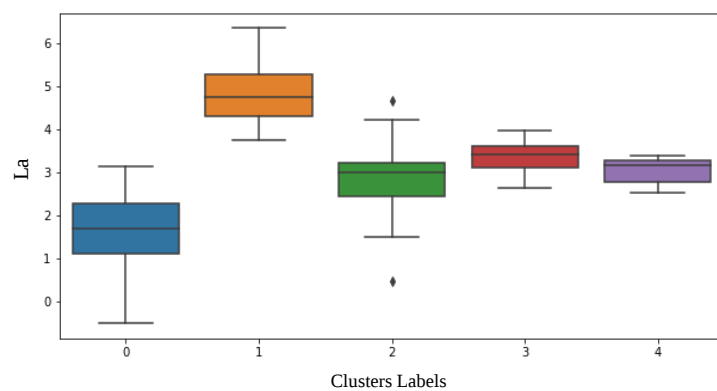
6.3.1 Análisis del Agrupamiento Jerárquico

A partir de los *boxplots* de la Figura 6.1 se destaca lo siguiente:

- AJC0 (rectángulos azules en Figura 6.1): engloba un amplio espectro de rocas (principalmente rocas volcánicas y brechas), presenta bajos valores de LREE (~25,26 ppm), HREE (~5,4 ppm) y Fe_2O_3 , altos valores de K_2O y valores intermedios de SiO_2 . En cuanto a alteraciones predominan feldespato potásico, cuarzo-sericita, calcita y clorita.
- AJC1 (rectángulos naranjos en Figura 6.1): conformada principalmente por brechas (tectónicas y volcánicas, además de brechas “verdes” según Rieger et al., 2010b), stockwork de especularita, y rocas volcánicas. Tiene predominancia de alteraciones potásica, clorita, sericita, cuarzo y clorita. Este clúster presenta altas concentraciones de LREE (~381,1 ppm), Fe_2O_3 y K_2O , y bajo contenido de HREE (~5,18 ppm), SiO_2 , Na_2O y CaO . Las altas concentraciones de Fe_2O_3 se pueden explicar por la presencia de especularita, clorita y sericita, así como también por la Dacita Meléndez y/o la presencia de magnetita. De la misma manera, altas concentraciones de K_2O se pueden atribuir a la presencia importante de alteración potásica, sericita, feldespato-potásico y biotita, la cual puede estar relacionada con la intensa alteración diseminada en la Brecha de magnetita en el distrito Candelaria-Punta del Cobre y/o a la alteración en la brecha de especularita en el distrito Mantoverde.
- AJC2 (rectángulos verdes en Figura 6.1): este clúster está compuesto principalmente por rocas volcánicas y/o hipabisales, diques dioríticos y fases de mineralización dacíticas. Este clúster presenta bajas concentraciones de REE (LREE ~88,98 ppm y HREE ~11 ppm). Contiene minerales de alteración predominantes de magnetita, actinolita, clorita, biotita y feldespato potásico. El bajo contenido de REE se puede deber al alto contenido de SiO_2 y las bajas concentraciones de Fe_2O_3 , Al_2O_3 y Na_2O ; aunque tiene un alto contenido de MgO .

Se cree que el alto contenido de SiO_2 se debe al tipo de rocas y a las alteraciones presentes.

- AJC3 (rectángulos rojos en Figura 6.1): compuesto por rocas volcánicas félsicas y volcanoclásticas. Alto contenido en REE (LREE $\sim 135,42$ ppm y HREE $\sim 68,71$ ppm), Na_2O y MgO ; y baja presencia de SiO_2 y K_2O . Los minerales de alteración predominantes son feldespato potásico, biotita, magnetita y clorita.
- AJC4 (rectángulos morados en Figura 6.1): conformada por rocas en su mayoría plutónicas. Presenta un alto contenido de Al_2O_3 , CaO , Na_2O , P_2O_5 , Y e Sr. Es el clúster con el segundo mayor contenido de HREE ($\sim 14,95$ ppm), aunque no es significativamente alto. Contiene alteración potásica, carbonitización y actinolita. El clúster presenta concentraciones altas de SiO_2 , P_2O_5 , Y y Sr; concentraciones intermedias de LREE ($\sim 120,93$ ppm), Fe_2O_3 , K_2O ; y bajo MgO . La concentración de REE lo vuelve el tercer clúster con mayor presencia. El contenido de REE puede estar ligado a la firma geoquímica intrusiva de las rocas que la conforman.



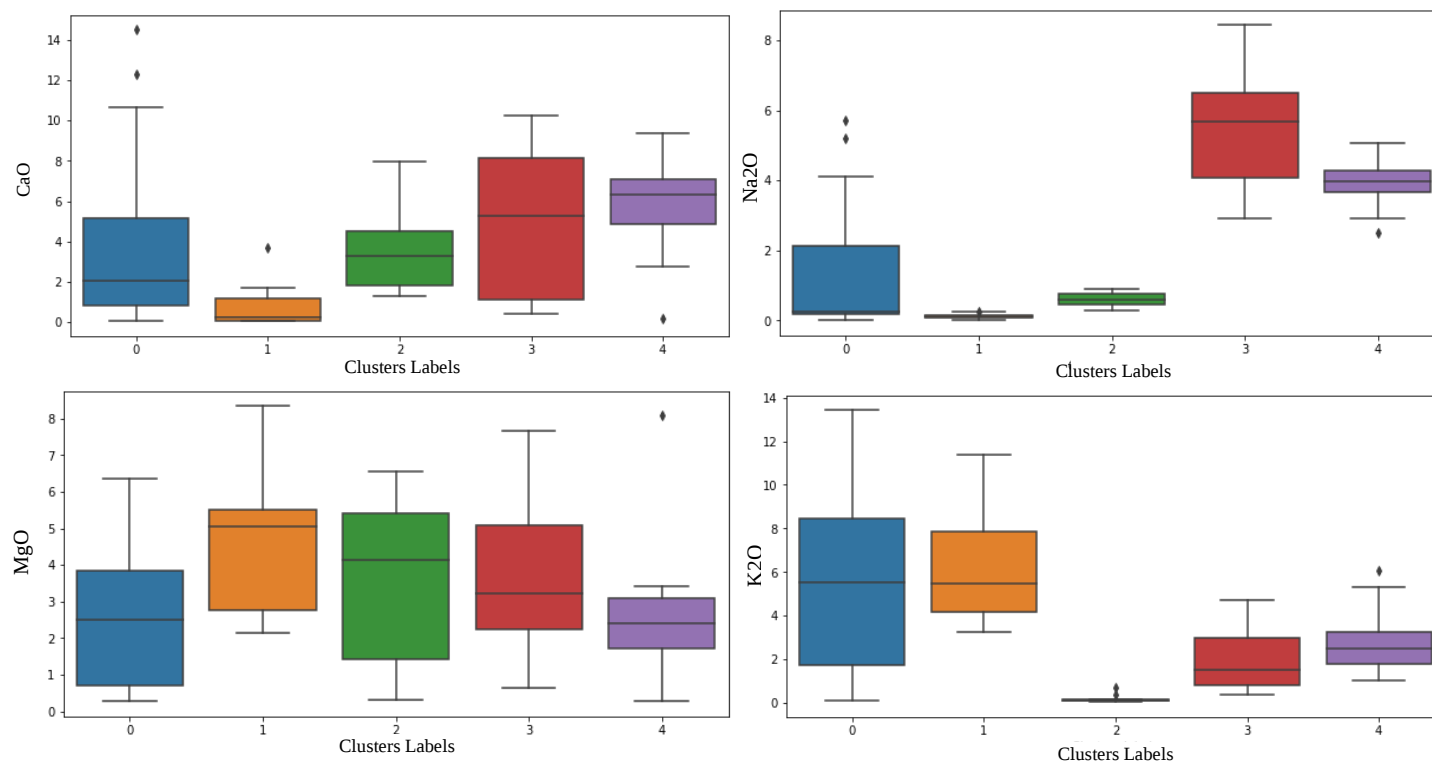


Figura 6.1. *Boxplots* de los clústeres obtenidos por Agrupamiento Jerárquico de las muestras. El eje vertical corresponde a los valores de ppm estandarizados mediante Standard Scaler.

Tabla 6.4 Características de los cinco clústeres obtenida mediante el análisis de Agrupamiento Jerárquico de las muestras.

Clúster	Numero de Muestras	Litología predominante	Alteración Predominante	Contenido
0	26	Brechas tectónicas, de especularita y verde; stockwork de especularita; Rocas volcánicas; Dacita Meléndez y Andesita Superior.	Kfps-Ser-qtz±cal; Kfps-chl-cal; Potásica; bio-act-Kfps	Bajos valores de LREE (5°) y HREE (4°); primero en K ₂ O; tercero en SiO ₂ ; cuarto en MgO, Fe ₂ O ₃ , CaO Y, P ₂ O ₅ e Sr; último en Na ₂ O.
1	11	Stockwork de especularita; Brechas tectónicas, de especularita y verde; Rocas volcánicas; Dacita Meléndez	Potásica; chl-ser±qtz; Kfps-ser-qtz±cal; ser-qtz±chl	Primero en LREE, Fe ₂ O ₃ y MgO; cuarto en Na ₂ O; último en HREE, SiO ₂ , CaO, Sr, P ₂ O ₅ e Y; segundo en K ₂ O.
2	13	Andesita Inferior; Dacita Meléndez; Andesita Superior; Diques de dacita Pre y Post-mineralización	Kfps -qtz-chl-bio; mt-act-bio-Kfps -qtz; Na-Ca-Fe(mt) K-Fe Ca-Fe	Cuarto en LREE; tercero en HREE, Y, P ₂ O ₅ y Sr; primero en SiO ₂ ; segundo en MgO; último en Fe ₂ O ₃ , Al ₂ O ₃ y K ₂ O.
3	13	Dacita Melendez; Andesita Superior; Rocas Volcanoclásticas; Rocas volcánicas.	Na-Ca-Fe(mt) k-Fe Ca-Fe; bio-qz-mt-kfps	Primero en HREE, Na ₂ O; segundo en LREE, Fe ₂ O ₃ , CaO, SiO ₂ , Al ₂ O ₃ e Y; tercero en, MgO, P ₂ O ₅ y Sr.
4	18	Rocas intrusivas; Diorita rica en piroxeno; Diorita rica en Hornblenda; Tonalita y Monzodiorita.	Kfps -bio-act-qtz; Potásica; bio- Kfps -act-qtz	Segundo en HREE, Na ₂ O y SiO ₂ , tercero en LREE, Fe ₂ O ₃ , y K ₂ O; primero en CaO, Y, P ₂ O ₅ , Al ₂ O ₃ y Sr; último en MgO.

Abreviaciones. Kfps: feldespato potásico; Ser: sericita; qtz: cuarzo; cal: calcita; bio: biotita; act: actinolita; chl: clorita; mt: magnetita

6.3.2. Análisis de *K-means*

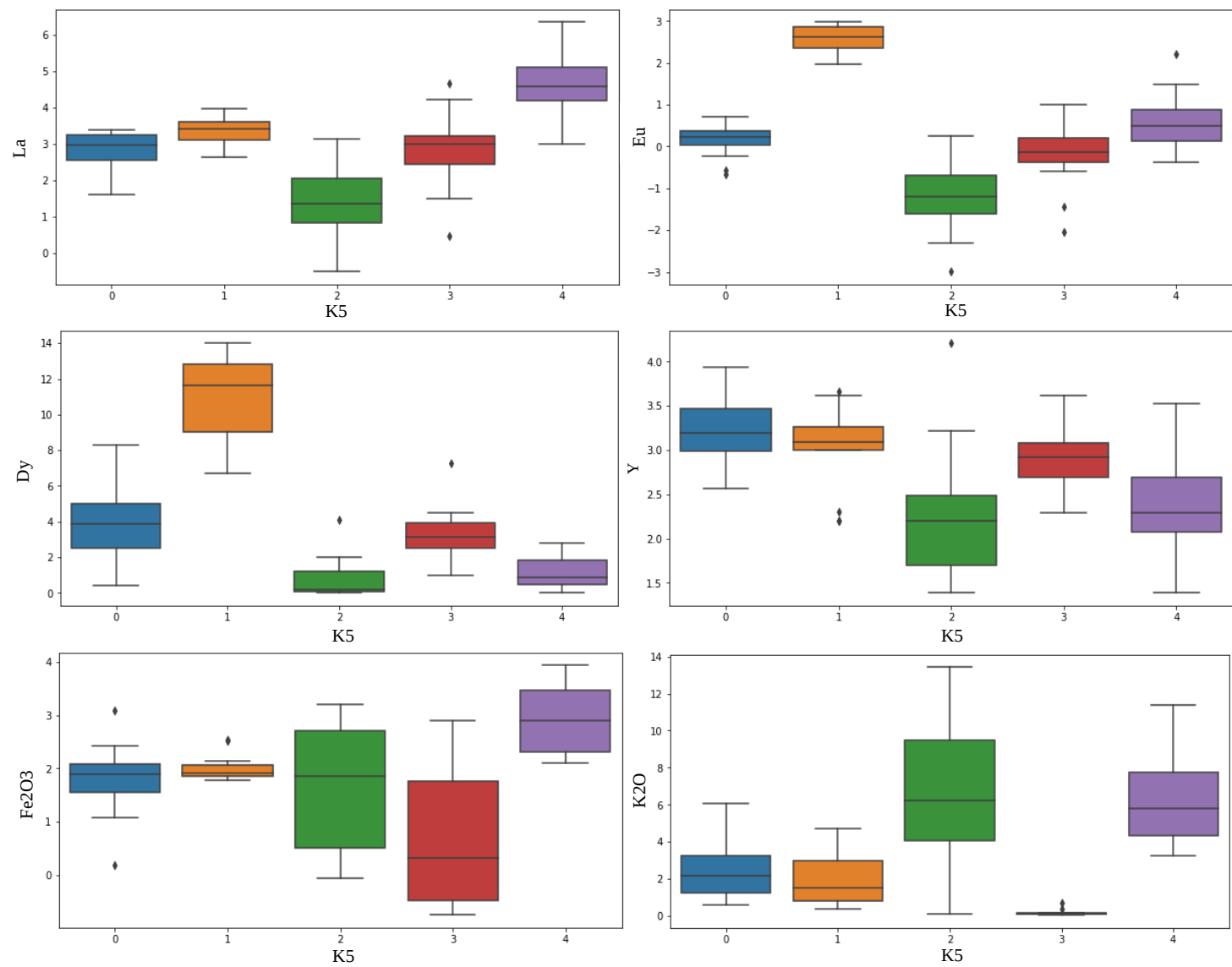
A partir de los *boxplots* de la Figura 6.2 se destaca lo siguiente:

- KMC0 está compuesto mayormente por rocas intrusivas como dioritas, monzodioritas, tonalitas. Es el tercer clúster con mayor presencia de REE (LREE ~99 ppm y HREE ~13,05 ppm). Es el clúster con mayor concentración de Y, Sr, P₂O₅, CaO e Al₂O₃, y el segundo en Na₂O y SiO₂. Tiene bajas concentraciones de MgO. En cuanto a las alteraciones tiene predominancia de potásica (feldespato potásico y biotita), actinolita, albita-epidota y granate. El alto contenido de Na₂O se puede explicar por las rocas y por la presencia de feldespato potásico, albita, y piroxeno. Para el contenido de Al se puede explicar mediante tipo de roca, clorita, feldespato, albita, epidota, biotita, granate.
- KMC1 está compuesto principalmente por rocas volcánicas como la Andesita Superior y Dacita Meléndez, además de rocas volcanoclásticas. Presenta las mayores concentraciones de REE (HREE ~135,42 ppm y LREE ~68,71 ppm), CaO y Na₂O. Además, tiene altos valores de Fe₂O₃, Y, TiO₂ Sr, P₂O₅ y Al₂O₃. En cuanto a las alteraciones, presenta predominancia de feldespato potásico-cuarzo-clorita-biotita, magnetita, y calco-sódica. Contiene cantidades intermedias de MgO, y bajas concentraciones de SiO₂ y K₂O. El alto contenido de CaO y Na₂O se puede explicar por la composición de las rocas volcánicas como también la presencia de alteración calco-sódica.
- KMC2 está conformado por rocas volcánicas, brechas y stockworks de especularita. Muestra las menores concentraciones de REE (LREE ~18,4 ppm y HREE ~4,3 ppm), TiO₂, MgO, Y y Sr. Contiene las mayores concentraciones de K₂O. Presenta concentraciones intermedias de SiO₂ y Al₂O₃, y bajo contenido de Fe₂O₃, Na₂O, CaO y P₂O₅. Tiene predominancia de alteración potásica (feldespato potásico), cuarzo, clorita, carbonitización y sericita. El bajo contenido de REE puede estar relacionada a la elevada presencia de K₂O y al bajo contenido de Fe₂O₃, Na₂O, CaO y MgO.
- KMC3 está constituido principalmente por rocas volcánicas como andesita. Contiene asociaciones de minerales de alteración como feldespato potásico-cuarzo, clorita, biotita, magnetita-actinolita-biotita-feldespato potásico y

alteración calco-sódica. En comparación con los demás clústeres, es el cuarto clúster en cuanto a presencia de REE (LREE ~88,98 ppm y HREE ~11,65 ppm). Contiene concentraciones intermedias de Na₂O, CaO, Y, P₂O₅ y Sr. Presenta la mayor concentración de SiO₂. Es el segundo en concentración de MgO. Es el clúster que presenta la menor concentración de Al₂O₃, Fe₂O₃ y K₂O. El bajo contenido de REE se puede relacionar con la baja presencia de Fe₂O₃, Na₂O, CaO y Al₂O₃; y a la alta presencia de SiO₂.

- KMC4 abarca principalmente brechas (volcánicas, tectónicas y “verde”), Stockwork de especularita y rocas volcánicas. Tiene una predominancia de alteración potásica y asociaciones de alteración como clorita, sericita-cuarzo, feldespato potásico-cuarzo-calcita y clorita-cuarzo-hematita. Presenta la mayor concentración de LREE (LREE ~343,35 ppm), Fe₂O₃ y MgO. Es el cuarto en cuanto a HREE (~5.52 ppm), Y, Sr y Al₂O₃. Presenta las menores concentraciones de SiO₂, CaO, Na₂O y P₂O₅. El alto contenido de LREE está en consonancia con el alto contenido de Fe₂O₃ y MgO, y con el bajo contenido de SiO₂.

El Anexo 5 presenta la concentración de cada REE en ppm con respecto a los valores de la corteza de la Tabla 2.1.



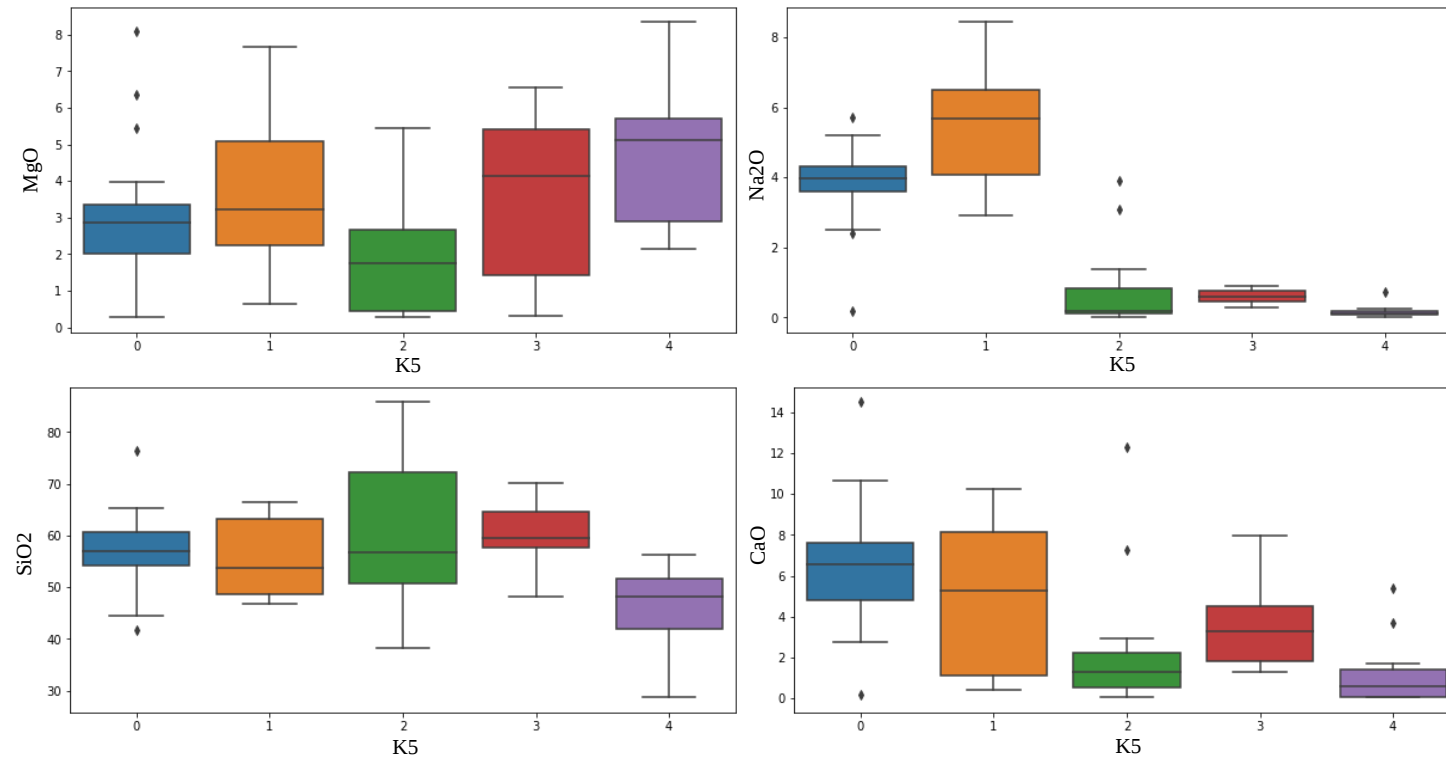


Figura 6.2. *Boxplots* de los clústeres obtenidos por *K-means* de las muestras. El eje vertical corresponde a los valores de ppm estandarizados mediante Standar Scaler.

Tabla 6.5. Características de los cinco clústeres obtenida mediante el análisis de *K-means*

Clúster	Número de Muestras	Litología predominante	Alteración Predominante	Contenido
0	24	Diorita rica en piroxeno; Diorita rica en hornblenda; Diorita rica en piroxeno; Monzodiorita; Tonalita; Cuarzo monzonita porfírica; Rocas plutónicas alteradas.	Cal-chl; ab-ep; Potásica; Carbonitización; bio-act-Kfps; Kfps -bio-act-qz; Gt-act- Kfps; bio- Kfps -act-qz	Segundo en HREE, SiO ₂ , Na ₂ O y CaO. tercero en LREE, K ₂ O y Fe ₂ O ₃ ; cuarto en MgO; primero en Y, Sr, Al ₂ O ₃ y P ₂ O ₅ ;
1	13	Andesita superior; Dacita Meléndez; rocas volcanoclásticas.	Kfps-qtz-chl-bio; Na-Ca-Fe(mg) k-Fe Ca-Fe; bio-qtz-mt-kfps; Potásica.	Primero en HREE y Na ₂ O; segundo en LREE, Fe ₂ O ₃ , CaO, Y, Sr, P ₂ O ₅ e Al ₂ O ₃ ; tercero en, MgO, P ₂ O ₅ y Sr; cuarto en SiO ₂ y K ₂ O.
2	19	Brecha volcánica, de especularita, tectónica y brecha verde; Stockwork especularita; Dacita Meléndez; rocas volcánicas.	Kfps-qtz±ser; potásica; Chl-ser±qtz; Kfps-chl-cal; Carbonitización; Kfs±ser-cal	Ultimo en REE, MgO, Y e Sr, primero en K ₂ O. tercero en SiO ₂ , Al ₂ O ₃ ; cuarto en Fe ₂ O ₃ , Na ₂ O, CaO y P ₂ O ₅ .
3	13	Andesita inferior; Dacita Meléndez; Andesita superior; Diques dacíticos pre y post mineralización; Dique Diorítico.	Kfps-qtz-chl-bio; mt-act-bit-kfdps-qtz; Na-Ca-Fe(mg) k-Fe Ca-Fe.	Tercero en HREE, CaO, Na ₂ O, Y P ₂ O ₅ y Sr; cuarto en LREE; primero en SiO ₂ . segundo en MgO; último en Al ₂ O ₃ , Fe ₂ O ₃ y K ₂ O.
4	12	Brecha volcánica, tectónica y verde; Manto Atacama; Dacita Meléndez; rocas volcanoclásticas; Stockwork de especularita.	Potásica; Chl-ser±qtz; Kfps-ser-qtz±cal; Chl-qtz-hm.	Primero en LREE, Fe ₂ O ₃ y MgO; cuarto en HREE, Al ₂ O ₃ , Y e Sr; segundo en K ₂ O; último en SiO ₂ , CaO, Na ₂ O y P ₂ O ₅ .

Abreviaciones. Kfps: feldespato potásico; Ser: sericita; qtz: cuarzo; cal: calcita; bio: biotita; act: actinolita; chl: clorita; mt: magnetita; ab: albita; ep: epidota; Gt: granate.

6.4 Correlación de resultados

6.4.1 Comparación Agrupamiento Jerárquico vs *K-means*

La Tabla 6.6 presenta las equivalencias de los clústeres obtenidos para los algoritmos. A partir de dicha tabla se puede observar que KMC1, KMC3 son equivalentes a AJC3 y AJC2 respectivamente. Por su parte KMC2 está conformada por 19 muestras que conforman a AJC0. A su vez, KMC0 está conformada por seis muestras que conforman AJC0 (21, 23, 25, 63, 71 y 72) y todas las muestras de AJC4, lo cual provoca una diferencia entre las características de KMC0 versus AJC0-AJC4. Finalmente, KMC4 está conformada por una muestra de AJC0 (2) y todas las muestras de AJC1.

Los clústeres que presentan una mayor presencia de REE son KMC1-AJC3 y KMC4-AJC1. KMC1-AJC3 contiene elevadas concentraciones de HREE, LREE, Na₂O, CaO, Y, Sr, P₂O₅ y Al₂O₃. KMC4-AJC1 presenta las mayores concentraciones de LREE, Fe₂O₃ y MgO, y las menores concentraciones de SiO₂. Las características que comparten ambos son la presencia de rocas volcánicas como Dacita Meléndez y rocas volcanoclásticas del Miembro Algarrobo, además de la presencia de minerales de alteración como cuarzo, clorita, y alteración potásica y calco-sódica.

Las menores concentraciones de REE se observan en el clúster KMC2-AJC0, el cual presenta las mayores concentraciones de K₂O y concentraciones intermedias de SiO₂. Además, presenta un bajo contenido en Al₂O₃, Na₂O, MgO, Y, Sr y P₂O₅. Los clústeres KMC0-AJC0-AJC4 presentan concentraciones bajas a intermedias de REE. Los tres clústeres anteriormente mencionados, están compuesto mayoritariamente por rocas intrusivas, lo que los diferencian de los otros clústeres que tienen una mayor presencia de rocas volcánicas y brechas.

Tabla 6.6 Comparación de clústeres para cada método

<i>K-means</i>	Agrupamiento Jerárquico
KMC0: Bajos (a intermedios) valores de HREE. Tercero en LREE. Baja presencia de SiO₂. Mayor concentración de Y e Al₂O₃ y segundo en Na₂O. Primero en P₂O₅. Cuarto en Fe₂O₃. Cuarto en K₂O (Bajo). Quinto en MgO. Tercero en TiO₂. Nulo MnO. Muestras: 21, 23, 25, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89 y 90.	AJC0: Bajos valores de LREE. Segundo en HREE. Primero en SiO₂. Segundo en Fe₂O₃. Cuarto en TiO₂. Muy bajo MnO. Alto MgO (2°). Cuarto en CaO. Muestras: 21, 23, 25, 71 y 72 AJC4: Contenido intermedio de LREE. Contenido intermedio de HREE (a bajo). Primero en Y. Tercero en TiO₂. Ultimo en MnO. Ultimo en MgO. Segundo en CaO. Primero en P₂O₅. Ultimo en Fe₂O₃. Tercero en SiO₂. Tercero en K₂O. Primero en Al₂O₃. Segundo en Na₂O. Muestras: 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89 y 90.
KMC1: Primero en HREE. Segundo en LREE. Primero en Na₂O y CaO. Tercero en SiO₂ y MgO. Cuarto en Fe₂O₃ y K₂O.	AJC3: Primero en HREE. Primero en Na₂O y CaO. Segundo en LREE, SiO₂ y Al₂O₃. Tercero en Fe₂O₃, P₂O₅ y MgO. Bajo contenido en K₂O
KMC2: Último en REE. Primero en SiO₂ y K₂O. Segundo en Fe₂O₃. Tercero en Na₂O. Bajo contenido de Al₂O₃, MgO, CaO, Y e Sr. Muestras: 5, 6, 7, 8, 12, 14, 15, 16, 17, 18, 19, 20, 24, 26, 65, 66, 68 y 69.	AJC0: Último en REE. Primero en K₂O. Segundo en Fe₂O₃. Bajo contenido en SiO₂, MgO, CaO. Tercero en Na₂O y Al₂O₃ Muestras: 5, 6, 7, 8, 12, 14, 15, 16, 17, 18, 19, 20, 24, 26, 65, 66, 68 y 69.
KMC3: Cuarto en REE. Segundo en SiO₂, MgO, P₂O₅ y Sr. Bajo contenido de Na₂O y CaO. Ultimo en K₂O, Fe₂O₃ y Al₂O₃.	AJC2: Cuarto en REE e Y. Primero en SiO₂. Segundo en MgO, P₂O₅ y Sr. Bajas concentraciones de CaO y Na₂O. Último en K₂O, Fe₂O₃ y Al₂O₃.
KMC4: Segundo en REE y K₂O. Primero en Fe₂O₃ y MgO. Cuarto en Al₂O₃. Bajo contenido en Y e Sr. ultimo en P₂O₅, Na₂O, CaO y SiO₂. Muestras: 0, 2, 3, 4, 9, 10, 11, 13, 22, 64, 67 y 70.	AJC1: Mayores concentraciones en REE. Último en Y. Primero en Fe₂O₃. Segundo en K₂O. Menores cantidades de P₂O₅, SiO₂ y Na₂O. Muestras: 0, 2, 3, 4, 9, 10, 11, 13, 64, 67 y 70. AJC0: Bajos valores de LREE. Segundo en HREE. Primero en SiO₂. Segundo en Fe₂O₃. Cuarto en TiO₂. Muy bajo MnO. Alto MgO (2°). Cuarto en CaO. Muestra: 22.

*Solo se presentan las muestras que conforman dos clústeres de AJ y uno de *K-means*.

6.4.2 Comparación entre PCA y matriz de correlación

La matriz de correlación (Figura 5.2) muestra que las HREE presentan una fuerte correlación entre sí y con Eu (LREE), y correlación moderada con Pr (LREE), Sm (LREE) y Na₂O. Por otro lado, las LREE presentan una fuerte correlación entre sí, y correlación moderada con Fe₂O₃ y MgO. Cabe también destacar que la matriz indica que las concentraciones de REE se correlacionan negativamente con K₂O y SiO₂. Los patrones de correlación antes descritos son en gran medida observados por el análisis de PCA. El primer componente principal (PC1) se correlaciona principalmente con las HREE, Pr, Sm y Eu, mientras que el segundo componente principal (PC2) se correlaciona principalmente con las LREE restantes y Fe₂O₃. Ambas técnicas proporcionan resultados similares y permiten identificar las afinidades de las REE con ciertos óxidos mayores.

6.4.3 Asociaciones de Elementos

En base la comparación de los resultados obtenidos con todos los análisis aplicados, se identifican las siguientes asociaciones de elementos que son relevantes para la exploración de REE:

- Las altas concentraciones de HREE se asocian con la alta concentración de Na₂O y CaO.
- El empobrecimiento de HREE coincide con el aumento de K₂O.
- Las altas concentraciones de LREE se asocian con la alta concentración de Fe₂O₃ y MgO.
- El empobrecimiento de LREE coincide con el aumento de SiO₂.
- Las muestras asociadas con alto contenido de LREE en el distrito Mantoverde, se encuentran cercanas a una falla con elongación N-S y se encuentran en unidades con mineralización económica, Brecha tectónica y Stockwork de especularita (Rieger et al., 2010a).
- Las muestras asociadas con alto contenido de REE (en especial LREE) en el distrito Candelaria-Punta del Cobre presentan litología de Dacita Meléndez y rocas volcánicas del Grupo Bandurrias.
- Los elementos Eu y Pr (LREE) muestran correlación con las HREE.

6.4.4 Agrupamiento utilizando dos clústeres para *K-means* y Agrupamiento Jerárquico

Como se discutió en la sección 6.1, adicionalmente se realizó Agrupamiento Jerárquico y *K-means* utilizando una agrupación en dos clústeres. Ambos algoritmos generaron dos grupos con exactamente las mismas características, y por lo tanto se realizó un único análisis de estos clústeres. El Clúster 0 (C0) contiene 14 muestras, las cuales se distribuyen casi en su totalidad en el distrito Candelaria-Punta del Cobre, y solo presenta una muestra (muestra 9) en Mantoverde. Por su parte el Clúster 1 (C1) incluye 67 muestras, la cual se distribuyó completamente en los distritos Candelaria-Punta del Cobre (57 muestras) y Mantoverde (24 muestras). Ver las Figuras 6.3 y 6.4 para los distritos Candelaria-Punta del Cobre y Mantoverde respectivamente. El Anexo 6 contiene los *boxplots* de estos dos clústeres. De estos resultados, se destaca que C1 presenta mayores concentraciones de REE, así como CaO, Na₂O, MgO, Y, Al₂O₃ y SiO₂. Por otro lado, C0 presenta mayores concentraciones de Fe₂O₃, K₂O, P₂O₅ y Sr. Las muestras en C1 exhibe las mismas litologías (Dacita Meléndez, Andesita superior y rocas volcánicas) a la de los clústeres con una alta presencia de REE, AJC1-KMC4 y AJC3-KMC1.

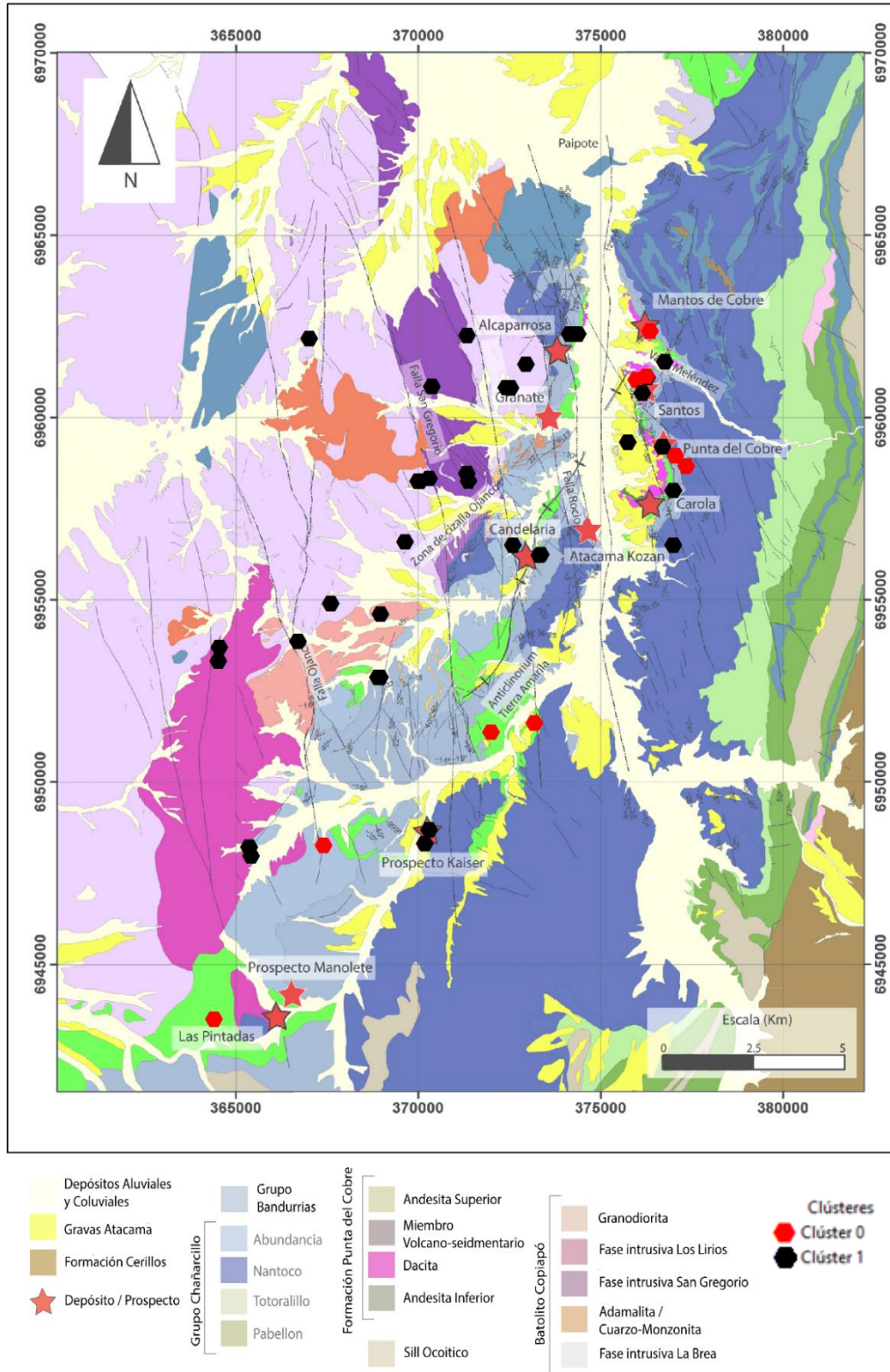


Figura 6.3. Distribución espacial de los clústeres obtenidos mediante el Agrupamiento Jerárquico y K-means en el distrito Candelaria-Punta del Cobre (geología tomada de Rieger et al., 2010a).

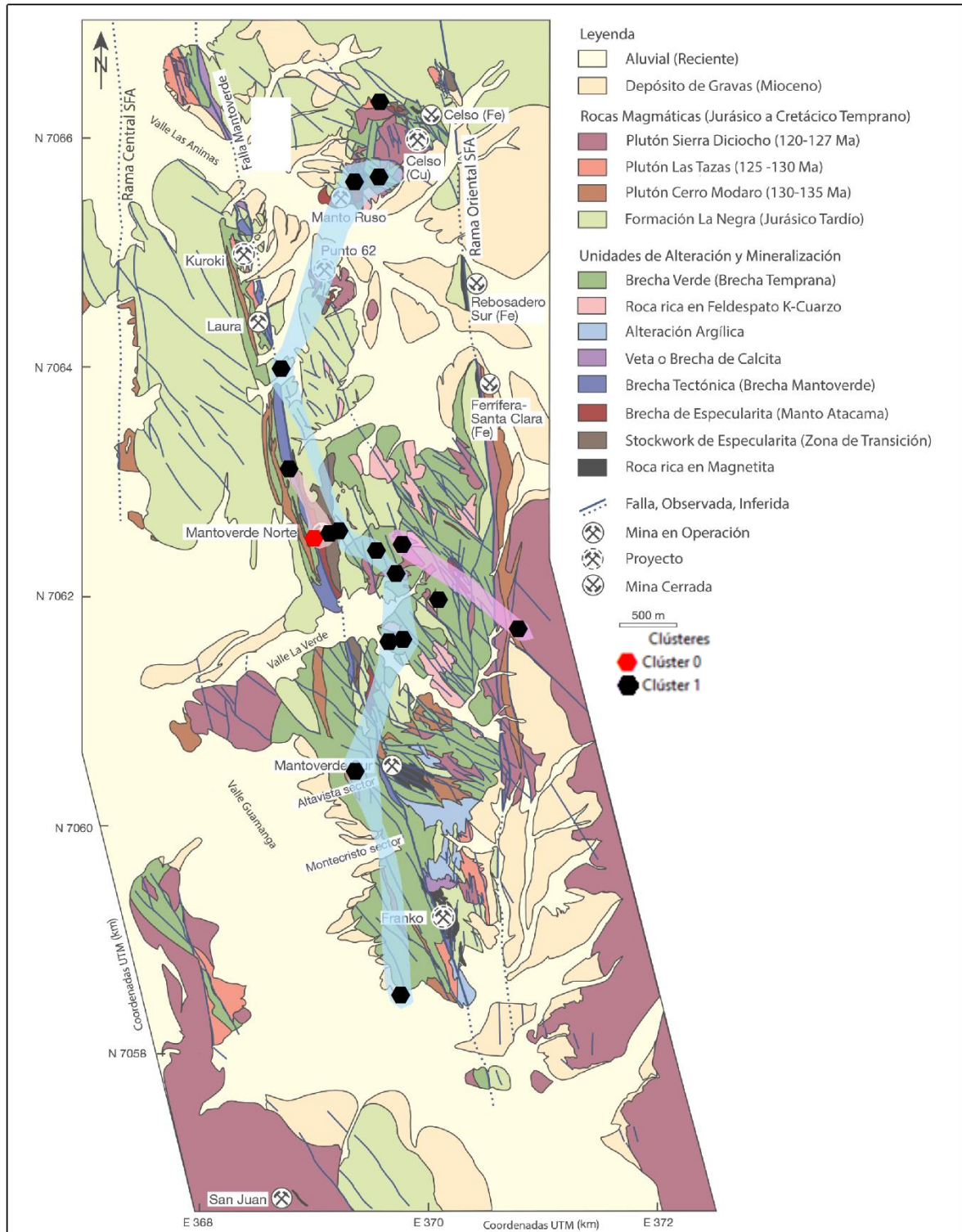


Figura 6.4. Distribución espacial de dos clústeres obtenidos mediante el Agrupamiento Jerárquico y *K-means* en el distrito Mantoverde (geología tomada de Rieger et al., 2010a).

6.5 Interpretación Geológica

La correlación positiva de las concentraciones de LREE con Fe_2O_3 , y a su vez la correlación negativa de los elementos anteriores con respecto a SiO_2 , se puede deber a que la presencia de Fe en los depósitos minerales disminuye con la presencia de Si y viceversa. Esta relación negativa se debe a que el Fe y el Si compiten por los mismos aniones y cationes en la solución, lo que puede afectar la formación y distribución de minerales en el depósito. Por ejemplo, la presencia de hierro puede disminuir la formación de minerales de silicato, mientras que la presencia de silicio puede disminuir la formación de minerales de hierro (Taylor y McLennan, 1985; Rudnick, 1995). La relación encontrada entre Ca y Na para las HREE es similar a lo reportado por Mirzababaei et al. (2021) en el depósito de hierro Se-Chahun en Irán. Manifestó que el metasomatismo dominante de Na-Ca que cambió débilmente la mineralogía del depósito, podría ser el responsable de la mineralización de REE en el depósito. La presencia de Fe, Ca, y Na puede ser indicativa de momentos tardíos en la formación del cuerpo intrusivo, mientras que la concentración de Al puede estar relacionada con estadios tempranos.

Los contenidos de Fe, Mg, Ca y Na muestran una correlación positiva con las concentraciones de REE, correlación cuya causa no es simple de determinar debido a la complejidad de los procesos geológicos involucrados. En general, se cree que la presencia de Fe y Mg puede estar relacionada con la formación de silicatos y fosfatos portadores de REE (Alibo y Nozaki, 1999; Bau, 1999; Franchin et al., 2005). Estos minerales pueden formarse a través de procesos de reemplazo de sulfuros y de recristalización de minerales preexistentes (Chen y Zhou, 2014, 2015). Por otro lado, la presencia de Ca y Na puede estar relacionada con la formación de carbonatos y sulfatos, respectivamente, que pueden ser portadores de REE (Anenburg et al., 2020; Carten, 1981). Sin embargo, es importante tener en cuenta que la relación de Fe, Mg, Ca y Na con la presencia de REE puede ser altamente dependiente del contexto geológico específico de cada depósito mineral. Por ejemplo, la composición química de las rocas encajantes, la presencia de fluidos hidrotermales y la historia tectónica pueden influir en la distribución y concentración de REE en un depósito mineral. El Fe y Mg en condiciones hidrotermales pueden estar presentes en la solución acuosa y reemplazar a otros elementos en la estructura cristalina de los minerales (Williams-Jones et al., 2012). Este proceso puede resultar en la formación

de minerales con una composición química diferente incluyendo la incorporación de REE en la estructura cristalina (Duan et al., 2023). Además, pueden afectar la solubilidad de los REE en la solución acuosa y su posterior precipitación en minerales (Lu et al., 2021).

La correlación positiva entre los contenidos de Pr (LREE) y las HREE se debe a la similitud en sus propiedades químicas y geoquímicas (Hai-qing et al., 2014). Ambas tierras raras tienen un comportamiento geoquímico similar en los procesos de mineralización y alteración, lo que les permite estar presentes juntas en los mismos minerales (Peng et al., 2018). Además, la presencia de Pr en los minerales también se relaciona con la presencia de HREE debido a la similitud en sus procesos de formación geológica (Han et al., 2019). Por ejemplo, ambas tierras raras pueden formarse a partir de la alteración de rocas ígneas o metamórficas y estar presentes en los mismos minerales asociados con estos procesos (Han et al., 2019).

En el distrito Mantoverde, Benavidez et al. (2007) reporta que la apatita y la allanita se formaron durante un evento hidrotermal que denominan “etapa de óxidos de hierro”. Esta es una etapa temprana que consiste en un evento geológico caracterizada por un intenso metasomatismo de hierro, con un dominio de hematita y magnetita (ver sección 3.2.2.1). La apatita y la allanita están relacionadas con la formación de magnetita (mt II) (Rieger et al., 2010a) durante la etapa de óxido de hierro, que se caracteriza por un intenso metasomatismo férrico asociado a una alteración potásica generalizada, silicificación y alteración hidrolítica de las rocas hospedantes. La etapa de óxidos de hierro es un período geológico clave en la formación de minerales de REE en el distrito Mantoverde, lo que la hace de interés para la exploración de estos elementos en el distrito Candelaria-Punta del Cobre. En este distrito también se ha descrito una etapa de óxido de hierro con una amplia presencia de magnetita-hematita, pero no son tan dominantes como en el caso del distrito Mantoverde. La fase de óxido de hierro del distrito Candelaria-Punta del Cobre presenta metasomatismo de Na temprano con una etapa principal de metasomatismo de Fe y alteración potásica, aunque los autores no indican que durante esta etapa se formaron minerales de REE de interés (Del Real et al., 2018, Marschiky Fontboté, 2001a). Sin embargo, es importante señalar que la presencia de apatita y allanita no garantiza la presencia de REE en cantidades económicamente viables. También se han encontrado concentraciones significativas de REE en las rocas huésped con patrones variados,

incluyendo enriquecimientos moderados de HREE o LREE y anomalías de Eu positivas o negativas (Rieger et al., 2010a, b). Los LREE se alojan en allanita y apatita en pequeñas cantidades. Se cree que la redistribución es el principal mecanismo de la concentración REE en el distrito Mantoverde, en lugar de agentes externos. Las rocas dioríticas alteradas en Mantoverde Norte y la brecha Manto Atacama de Mantoverde Norte también presentan concentraciones significativas de REE (Rieger et al., 2010a, b).

La presencia de REE en el distrito Candelaria-Punta del Cobre estaría relacionada, al menos en parte, a la presencia de allanita (Marschik et al. 2000, 2003; C. Osterman en Ray et al., 2002; Del Real et al., 2018). En el distrito Candelaria-Punta del Cobre, la allanita y el apatito se han identificado como fuentes de los REE en los depósitos de Candelaria, Carola y Socavón Rampa, pero ambos minerales sólo están presentes en pequeñas cantidades (Marschik et al., 2000; Marschik y Fontbote, 2001b; Hopf, 1990; Ryan et al., 1995). En Candelaria se encuentran localmente pirrotita, esfalerita, elevadas concentraciones de elementos ligeros de tierras raras (LREE) y trazas de molibdenita y arsenopirita (Mathur et al., 2002). La mineralización de Candelaria es anómala en LREE, Zn, Mo y Ag. Las concentraciones elevadas de tierras raras y Zn se dan principalmente cerca del contacto del Miembro Geraldo-Negro con el Miembro Algarrobos, y también cerca del contacto inferior de las rocas tobáceas que se correlacionan con la "Trinidad Siltstone", lo que indica un control estratigráfico sobre la deposición de sus minerales hospedantes (Marschik et al., 2000). Del Real et al. (2018) reporta que la Andesita Inferior muestra dos grupos geoquímicos distintivos dentro del miembro. El primer grupo está enriquecido en LREE, tiene una anomalía negativa de Eu y está empobrecido en MREE. Estas muestras fueron tomadas de niveles someros del miembro en el yacimiento de Alcaparrosa y de profundos sondeos de exploración del distrito al sur del yacimiento de Candelaria, lo que sugiere que este grupo no está restringido espacial o estratigráficamente en la Andesita Inferior. Además, las rocas monzodioríticas y monzoníticas cuarzosas del mismo distrito también presentan contenidos ligeramente elevados de LREE (Marschik et al., 2000, 2003).

Debido a la superposición de diferentes tipos de alteración no es posible reconocer un tipo de alteración específico a partir de los datos de las muestras. La alteración hidrotermal

puede agregar elementos y, en consecuencia, las agrupaciones podrían separarse por el grado de alteración en lugar de las diferencias en la geoquímica primaria. Las rocas han experimentado una intensa alteración, con la presencia de alteración potásica y clorita (Rieger et al., 2010a; Del Real et al., 2018; Marschik et al., 2000). La alteración cálcico-hierro-potásica es común en los clústeres con mayor contenido de REE, tal como se ha documentado por Montreuil et al. (2013, 2016), aunque su papel en el enriquecimiento de REE no está claramente determinado.

La alteración tipo férrica puede precipitar minerales como la apatita y la allanita. En el distrito Mantoverde, por ejemplo, se ha demostrado que la allanita y la apatita se formaron durante la etapa en que precipitan los óxidos de hierro (Rieger et al., 2010a; Ruiz et al., 2020). Por su parte, la alteración propilítica podría estar relacionada a la generación del fluorapatito. El distrito Candelaria-Punta del Cobre muestra el mismo tipo de alteraciones, y también se ha reportado la presencia de apatita y allanita (Marschik et al., 2000; Hopf, 1990; Ryan et al., 1995). Esta relación entre la alteración y mineralización en depósitos IOCG y la presencia de minerales portadores de REE ha sido documentada en el depósito de Olympic Dam, donde se ha identificado la presencia de minerales como la apatita y el fluorapatito en asociación con la mineralización de cobre y oro (Krnetá et al., 2018; Apukhtina et al., 2017). Sin embargo, es importante realizar un estudio detallado de las condiciones geológicas específicas en los distritos mineral para poder entender la influencia de Fe, Mg, Ca, Na, Si y K; el rol de las alteraciones y etapas de mineralización en la presencia de REE.

Sarparandeh y Hezarkhani (2016), en el depósito Se-Chahun, reportó que los minerales de hierro fosforado tienden a ser más propenso a albergar altas concentraciones de REE y contener apatita y monacita. También reporta la presencia de fluorapatita como principal mineral de REE. En otro trabajo, Zaremotlagh et al. (2016) propone que la mineralización de REE en el depósito Choghart podría estar controlada por mineralización de fosfato. Los fluidos implicados en el transporte y deposición de REE suelen tener una alta actividad de cloruro y posiblemente de fluoruro (Banks et al., 1994). Otros autores sugieren que los complejos de fluoruro más que controlar el transporte de REE en los sistemas hidrotermales, más bien actúan como ligando que promueve la precipitación de

minerales de REE como la fluocerita y bastnäsita (Williams-Jones et al., 2012). Además, estudios han demostrado que las alteraciones hidrotermales asociadas con intrusivos ígneos, como los intrusivos alcalinos y graníticos, pueden ser fuentes importantes de REE (Williams et al., 2005).

La relación entre P y REE ha sido ampliamente documentada (Bau y Dulski, 1996; Mavrogenes y O'Neill 1999; Jiyan et al., 2013), sin embargo, a pesar de estas correlaciones, en este trabajo no se ha encontrado una correlación evidente entre los contenidos de P_2O_5 y LREE. Emsbo et al. (2015) estudiaron los depósitos de fosfato sedimentario de la cuenca del río Yangtze en China y encontraron que no había correlación entre los contenidos de P_2O_5 y los contenidos de LREE, indicando que "el enriquecimiento de LREE en los depósitos de fosfato no está relacionado con el contenido de P_2O_5 , sino con la absorción y coprecipitación de LREE con minerales fosfáticos en condiciones especiales". Es importante tener en cuenta que la relación entre P y LREE puede ser compleja y depender de muchos factores, como la mineralogía de los depósitos y las condiciones geológicas en las que se formaron.

A pesar de no reportarse otros minerales de REE en los distritos estudiados, la presencia de REE podría ocurrir como impurezas en otros minerales. Morales (2022) observó anomalías de Ce en muestras de magnetita en los depósitos Punta del Cobre (2400 ppm) y Granate (2160 ppm). Sin embargo, estas anomalías son muy puntuales y no se pudo establecer una relación espacial respecto a su concentración debido a la falta de muestras en las que se detectó este elemento. Solís (2022) en el distrito Punta del Cobre reportó concentraciones anómalas de Ce y La, aunque de manera localizada y en la mayoría de los casos, las cantidades presentes son muy bajas o casi nulas. Además, observó que el evento de feldespato potásico-calcopirita posee altas concentraciones de ambos elementos, aunque solo analizó una muestra con la ocurrencia bien definida de este evento y no cuenta con suficiente información para interpretar la asociación mineralógica específica en relación con la presencia de La y Ce, por lo que no es concluyente. También se han reportado la presencia de REE (La) en los cristales de pirita en el distrito Candelaria-Punta del Cobre (Del Real et al., 2018, 2020; Velásquez et al., 2020).

En Olympic Dam, Kontonikas et al. (2018) sugiere que los feldespatos primarios como ortoclasa y la plagioclasa contienen REE en su estructura cristalina, pero no contribuyen significativamente a la mineralización de tierras raras en los sistemas IOCG estudiados. En cambio, los feldespatos secundarios como la sericita y la albita muestran una movilidad significativa de las REE debido a la entrada de fluidos hidrotermales ricos en tierras raras en la estructura cristalina de los feldespatos secundarios. Durante la etapa de oxidación, se liberan las REE almacenadas en los feldespatos secundarios y se depositan como óxidos de tierras raras. Los autores sugieren que la liberación de las REE de los feldespatos puede ocurrir mediante procesos de disolución-precipitación, reemplazo mineral, o una combinación de ambos. Además, también se menciona que la formación de las mineralizaciones de tierras raras en los sistemas IOCG puede estar relacionada con procesos de oxidación y precipitación de óxidos e hidróxidos de hierro y cobre, que pueden atrapar y concentrar las REE. Schmandt et al. (2019) reportó que la monacita y la xenotima son las principales fases de fosfatos de REE presentes en las etapas tempranas de mineralización de IOCG en Olympic Dam asociadas con la mineralización de cobre temprana y se encuentran en las rocas ígneas y metamórficas que rodean al yacimiento. Por otro lado, la florencita es una fase de fosfato de REE que se encuentra en las etapas posteriores de mineralización en Olympic Dam, asociada con la mineralización de uranio-oro. Esta fase se encuentra en las vetas hidrotermales de cuarzo que atraviesan la roca huésped y está relacionada con la actividad hidrotermal tardía en el yacimiento.

Lo expuesto anteriormente guarda similitud con lo observado en los distritos estudiados. En el distrito Mantoverde se observa un factor de interés la etapa de óxidos de hierro, en la cual se reporta que en esta etapa se formaron la apatita y allanita, además de hallarse alta presencia de Fe y S, por lo que las REE se podrían encontrar en forma de impurezas en minerales conformados por Fe y Si.

6.6 Implicancias para la exploración

Los resultados obtenidos en esta tesis sugieren que las HREE tienden a agruparse entre sí y con Na_2O . Esta asociación de elementos podría indicar una asociación de los HREE con minerales de alteración con alta presencia de Na, posiblemente asociados a eventos de alteración calco-sódica. Por otro lado, las LREE aparecen correlacionadas con Fe_2O_3 y

MgO. Además, se ha observado que, a pesar de ser una LREE, el Pr presenta una fuerte correlación con los HREE. Por su parte, el Sm y el Eu, que también son LREE, tienen una buena correlación con ambos subtipos de REE y con Na₂O e Y. La alteración férrica podría desempeñar un rol en el enriquecimiento o removilización de LREE (Williams et al., 2005).

En el distrito Candelaria-Punta del Cobre, ninguno de los autores que han reportado la presencia de REE (Marschik et al., 2000; Marschik y Fontbote, 2001b, Del Real et al., 2018) ha profundizado en los procesos de mineralización responsables de REE como tampoco ha identificado algún evento de mineralización o alteración relacionado. Aunque se ha reportado la presencia de allanita y apatita en este distrito, los contenidos de estos minerales son escasos y por lo tanto se requiere más investigación para comprender los procesos de mineralización de REE en este distrito.

En el distrito Mantoverde también se ha observado la presencia de minerales portadores de REE como la allanita y apatita (Rieger et al., 2010a, b). También, se ha observado la presencia de fluorapatita (Benavides et al., 2007), pero aún se desconoce su contribución en términos de REE. Las muestras 0, 2, 4, 10 y 11 del clúster KMC4-AJC0-AJC1 se encuentran en la zona de transición y de stockwork de especularita, en las cuales se haya la presencia de apatito (Rieger et al., 2010a, b). La formación de estos minerales portadores de REE se produciría en la parte final del evento de Óxidos de Hierro interpretado por Benavides et al. (2007), por lo cual este evento podría ser de interés. La comprensión completa de los procesos geológicos que llevaron a la formación de estos minerales es crucial para evaluar el potencial de la región como fuente de REE.

La alteración calco-sódica podría jugar un rol en la concentración y/o removilización de HREE, de la misma manera la alteración férrica para las LREE. A partir de lo expuesto anteriormente, se proponen considerar los siguientes puntos para la exploración de REE en los distritos estudiados:

1. La allanita y apatita, minerales portadores de REE, pueden estar presentes en las etapas tempranas de mineralización en las rocas volcánicas en los distritos. En el distrito Mantoverde, estos minerales portadores de REE se asociarían también a la parte final del evento de Óxidos de Hierro.

2. En el distrito Candelaria-Punta del Cobre las REE podrían estar presentes como impurezas en magnetita, pirita, sericita y feldespatos secundarios como la albita.
3. Las HREE podrían estar relacionadas con minerales de alteración con alta presencia de Na, posiblemente asociados a eventos de alteración calco-sódica.
4. Analizar la correlación entre los contenidos de Na_2O , CaO y HREE. El CaO puede desempeñar un papel en la distribución y concentración de HREE en los depósitos de REE.

La exploración de tierras raras en depósitos IOCG requiere un enfoque integral y multidisciplinario para identificar áreas potenciales de mineralización. La relación de Fe_2O_3 y MgO con LREE, y de Na_2O y CaO con HREE, puede ser un indicador importante para guiar la búsqueda de estos elementos críticos. Además de considerar estas relaciones, es importante tener en cuenta otros indicadores geológicos, como la alteración hidrotermal (calco-sódica, propilítica, férrica, etc.), la distribución de los elementos de traza, etc. En resumen, el uso de la relación entre Fe_2O_3 y MgO con las LREE, y entre Na_2O y CaO con las HREE, junto con un enfoque multidisciplinario, puede ser una guía efectiva para la exploración de REE en los distritos de IOCG en el área de estudio.

CAPÍTULO 7: CONCLUSIÓN

En esta tesis se evaluó la aplicabilidad y la efectividad de la aplicación de dos algoritmos de *Machine Learning* no supervisado, Agrupamiento Jerárquico y *K-means*, en el estudio de REE en distritos con mineralización tipo IOCG. Estas metodologías proporcionan un nuevo enfoque para identificar patrones de correlación entre las concentraciones de REE y otros elementos químicos o minerales, patrones que pueden ser utilizados como guías para la exploración de mineralización de REE. Estos algoritmos, junto al análisis de la matriz de correlación y el análisis de componentes principales, fueron aplicados en una base de datos de 81 muestras tomadas desde los distritos Candelaria-Punta del Cobre y Mantoverde.

A partir de esta investigación, se concluye lo siguiente:

- a) Los resultados obtenidos a través de la aplicación de los algoritmos de *K-means* y Agrupamiento Jerárquico son consistentes entre sí. Ambos algoritmos obtuvieron grupos o clústeres de muestras con características similares, y las diferencias encontradas entre ellos fueron mínimas. Estos resultados sugieren que ambas metodologías son efectivas y pueden ser utilizadas de manera complementaria para el análisis de patrones geoquímicos en la exploración de yacimientos de REE.
- b) Los resultados obtenidos por los algoritmos son consistentes con los resultados obtenidos mediante el Análisis de Componentes Principales y la matriz de correlación.
- c) Ambos algoritmos agruparon: (1) a todas las HREE y algunas LREE (Pr, Sm, Eu) con Na_2O ; (2) las LREE (La, Ce, Nd) junto con Fe_2O_3 y MgO , y en menor medida Al_2O_3 .
- d) Las concentraciones de REE disminuye con el aumento de los contenidos de SiO_2 y K_2O .
- e) Los resultados del análisis de Agrupamiento Jerárquico y *K-means* muestran diferencias en la composición de los clústeres con mayores contenidos de REE entre los distritos Mantoverde y Candelaria-Punta del Cobre. Mientras que en Mantoverde las muestras se ubican en brechas tectónicas y stockwork de

especularita, en Candelaria-Punta del Cobre las muestras se encuentran principalmente en rocas volcánicas como la Dacita Meléndez y rocas volcanoclásticas.

- f) Los minerales que podrían albergar la mineralización de REE varían dependiendo del depósito. La allanita y el apatito son los únicos minerales portadores de REE registrados en ambos distritos estudiados, aunque en pequeñas cantidades. Las REE también pueden presentarse como impurezas en otros minerales, como en magnetita, pirita, sericita y en feldespatos secundarios como la albita.
- g) En el distrito Mantoverde se observa un intenso metasomatismo férrico, silicificación y alteración calco-sódica de las rocas hospedantes durante la etapa de mineralización óxidos de hierro temprana, lo que sugiere que estas asociaciones podrían servir como indicadores útiles en la exploración de REE. Por su parte el distrito Candelaria-Punta del Cobre también presenta una fase de óxidos de hierro temprana, aunque no presenta evidencia del desarrollo de allanita y apatito durante esta etapa.

Al utilizar Agrupamiento Jerárquico y *K-means* para explorar REE, es importante seleccionar adecuadamente las variables y preprocesar los datos. El número de clústeres debe ser seleccionado cuidadosamente para evitar resultados poco útiles. La interpretación de los resultados debe ser realizada cuidadosamente para evitar conclusiones erróneas y se debe combinar con otras técnicas estadísticas y de visualización. La metodología empleada en este trabajo puede ser útil para explorar la distribución de REE y para identificar áreas potenciales de mineralización. Estas técnicas pueden mejorar la eficiencia en la exploración y reducir los costos, pero es importante tener en cuenta las limitaciones y precaución al interpretar los resultados. No entregan información sobre la génesis de las rocas o la mineralización y no reemplazan la necesidad de llevar a cabo análisis geológicos y geoquímicos detallados.

Como forma de profundizar en el estudio de REE se proponen las siguientes recomendaciones:

- Emplear clústeres obtenidos con Agrupamiento Jerárquico y *K-means* en algoritmos de aprendizaje supervisado para clasificar muestras con potencial en REE.
- Ampliar la base de datos agregando parámetros adicionales como la profundidad de la muestra y otros elementos químicos con importancia económica. La profundidad es importante porque las muestras tomadas a diferentes profundidades pueden contener diferentes tipos y concentraciones de REE.
- Realizar un análisis detallado de las alteraciones y su relación con el enriquecimiento o empobrecimiento de REE.
- Identificar la correlación de LREE con Fe_2O_3 y MgO , y la fuerte correlación del Pr con los HREE, y de Sm y Eu con ambos subtipos de REE y con Na_2O e Y.
- Aplicar este análisis a conjunto de datos obtenidos de depósitos genética y espacialmente asociados (e.g. depósitos tipo IOA).
- Realizar más estudios para entender la génesis y evolución de la mineralización de REE y su relación con las rocas y estructuras geológicas presentes.

REFERENCIAS

- Abdolmaleki, M., Rasmussen, T. M., & Pal, M. K. (2020). Exploration of iocg mineralizations using integration of space-borne remote sensing data with airborne geophysical data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 43(B3), 9–16. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-9-2020>
- Abraitis, P. K., Patrick, R. A. D., & Vaughan, D. J. (2004). Variations in the compositional, textural and electrical properties of natural pyrite: a review. *International Journal of Mineral Processing*, 74(1-4), 41-59.
- Acock, A. C., & Stavig, G. R. (1979). A measure of association for nonparametric statistics. *Social Forces*, 57(4), 1381-1386.
- Aguilef, S. G. (2012). Analisis geológico de estudios geofísicos de Polarización Inducida (IP) en sistemas IOCG y pórfido cuprífero del Norte de Chile. Tesis, Universidad de Chile.
- Aguilera, C. I. (2000). Un enfoque gerencial de la teoría de las restricciones. *Estudios gerenciales*, (77), 53-69.
- Aitchison, J. (1986). Monographs on statistics and applied probability. The statistical analysis of compositional data.
- Aizerman, M. A., Braverman, E. M., & Rozonoer, L.I. (1964a). Theoretical foundations of the potential function method in the problema of training automata to classify input situations, *Automat. Remote Contr. (USSR)*25(6).
- Aizerman, M.A., Braverman, E.M., & Rozonocr, L.I. (1964b) The probabilistic problema of training automata to recognize patterns and the potential function method, *Automat. Remote Contr. (USSR)*25(9).
- Alberca, A. S. (2022a, mayo). La librería Numpy. Aprende con Alf. Recuperado 24 de agosto de 2022, de <https://aprendeconalf.es/docencia/python/manual/numpy/>
- Alberca, A. S. (2022b, mayo). La librería Pandas. Aprende con Alf. Recuperado 24 de agosto de 2022, <https://aprendeconalf.es/docencia/python/manual/pandas/>

- Alibo, D. S., & Nozaki, Y. (1999). Rare earth elements in seawater: particle association, shale-normalization, and Ce oxidation. *Geochimica et cosmochimica acta*, 63(3-4), 363-
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Álvarez, S. A. (2021). Desarrollo de una aplicación para visualización, creación y modificación de mapas paleogeográficos. [Título Profesional, Universidad de Chile]. Repositorio Académico de la Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/182285>
- Amat J. (2020, octubre). Machine Learning con Python y Scikitlearn. Ciencia de datos. Recuperado 23 de agosto de 2022, de https://www.cienciadedatos.net/documentos/py06_machine_learning_python_scikitlearn.html
- Amigó, E., Gonzalo, J., Artiles, J., & Verdejo, F. (2009). A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information retrieval*, 12, 461-486.
- Andrews, J. L. (2018). Addressing overfitting and underfitting in Gaussian model-based clustering. *Computational Statistics & Data Analysis*, 127, 160–171. doi:10.1016/j.csda.2018.05.015.
- Anenburg, M., Mavrogenes, J. A., Frigo, C., & Wall, F. (2020). Rare earth element mobility in and around carbonatites controlled by sodium, potassium, and silica. *Science Advances*, 6(41), eabb6570.
- Anthony, J. W., Bideaux, R. A., Bladh, K. W., & Nichols, M. C. (2001). *Handbook of mineralogy* (Vol. 1). Mineral Data Publ.
- Ao, S. I., Rieger, B. B., & Amouzegar, M. (Eds.). (2010). *Machine learning and systems engineering* (Vol. 68). Springer Science & Business Media.
- Apolloni, B., Ghosh, A., Alpaslan, F., & Patnaik, S. (Eds.). (2005). *Machine learning and robot perception* (Vol. 7). Springer Science & Business Media.
- Apukhtina, O. B., Kamenetsky, V. S., Ehrig, K., Kamenetsky, M. B., Maas, R., Thompson, J., McPhie, J., Cibanu, C. L., & Cook, N. J. (2017). Early, deep magnetite-fluorapatite mineralization at the olympic dam Cu-U-Au-Ag deposit, South Australia. *Economic Geology*, 112(6), 1531-1542.

- Arabasz, W. (1971). Geological and geophysical studies of the Atacama fault zone in northern Chile: Unpublished. PhD. thesis. California Institute of Technology, Pasadena, USA.
- Arévalo, C. (1999). The Coastal Cordillera/Precordillera boundary in the Tierra Amarilla area (27° 20' 27 40'S/70° 05'-70 20'W), northern Chile, and the structural setting of the Candelaria Cu-Au ore deposit: Unpublished Ph.D. thesis, Kingston, Canada, University of Queens.
- Arévalo, C. (2005). Carta Copiapó, Región de Atacama. Servicio Nacional de Geología y Minería, Carta Geológica de Chile, Serie Geología Básica, 91, 54.
- Arévalo, C., Grocott, J., & Welkner, D. (2003). The Atacama fault system in the Huasco province, southern Atacama desert, Chile. In Congreso Geológico Chileno (Vol. 10).
- Arévalo, C., Grocott, J., Martin, W., Pringle, M., & Taylor, G. (2006). Structural Setting of the Candelaria Fe Oxide Cu-Au Deposit, Chilean Andes (27° 30' S). *Economic Geology*, 101(4), 819-841.
- Argall, G.O., (1980). Three orebodies of Bayan Obo. *World Mining*, 33(1): 38-41.
- Ari, N., & Ustazhanov, M. (2014). Matplotlib in python. In 2014 11th International Conference on Electronics, Computer and Computation (ICECCO) (pp. 1-6). IEEE.
- Atherton, M.P., & Aguirre, L. (1992). Thermal and geotectonic setting of Cretaceous volcanic rocks near Ica, Peru, in relation to Andean crustal thinning. *J. S. Am. Earth Sci.* 5, 47–69.
- Atwood, D. A., (2013). *The rare earth elements: fundamentals and applications*. (1st ed.). John Wiley & Sons.
- Avendaño, P. J. (2017). Evaluación de factibilidad técnico-ambiental de una planta de extracción de tierras raras en Chile. [Título Profesional, Universidad de Chile]. Repositorio Académico de la Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/150686>
- AWS Amazon. (s. f.). *Entrenamiento de modelos de ML - Amazon Machine Learning*. AWS Amazon. Recuperado 23 de marzo de 2022, de https://docs.aws.amazon.com/es_es/machine-learning/latest/dg/training-ml-models.html
- Babaie, H., & Davarpanah, A. (2020). A classification and predictive model of the complex REE mineral system. In EGU General Assembly Conference Abstracts (p. 6324).

- Baioumy, H. M., Ahmed, A. H., & Khedr, M. Z. (2014). A mixed hydrogenous and hydrothermal origin of the Bahariya iron ores, Egypt: evidences from the trace and rare earth element geochemistry. *Journal of Geochemical Exploration*, 146, 149-162.
- Balcerzak, M. (2000). Separation, Preconcentration and Spectrophotometry in Inorganic Analysis. *Analytical Spectroscopy Library*, Volume 10. Elsevier Science & Technology.
- Ball, N. M., & Brunner, R. J. (2010). Data mining and machine learning in astronomy. *International Journal of Modern Physics D*, 19(07), 1049-1106.
- Banks, D. A., Yardley, B. W. D., Campbell, A. R., & Jarvis, K. E. (1994). REE composition of an aqueous magmatic fluid: a fluid inclusion study from the Capitan Pluton, New Mexico, USA. *Chemical Geology*, 113(3-4), 259-272.
- Bao, Z., & Zhao, Z. (2008). Geochemistry of mineralization with exchangeable REY in the weathering crusts of granitic rocks in South China. *Ore Geology Reviews*, 33(3-4), 519-535.
- Barlow, H. B. (1989). Unsupervised learning. *Neural computation*, 1(3), 295-311.
- Barra, F., Reich, M., Selby, D., Rojas, P., Simon, A., Salazar, E., & Palma, G. (2017). Unraveling the origin of the Andean IOCG clan: A Re-Os isotope approach. *Ore Geology Reviews*, 81, 62-78.
- Barrett, P., Hunter, J., Miller, J. T., Hsu, J. C., & Greenfield, P. (2005). matplotlib--A Portable Python Plotting Package. In *Astronomical data analysis software and systems XIV* (Vol. 347, p. 91).
- Barton M. D. (2014). Iron Oxide (-Cu-Au-REE-P-Ag-U-Co) Systems. In *Treatise of Geochemistry* eds. H. Holland and K. Turekian) vol. 13, pp. 515-536.
- Barton M. D. (2015). Arc-scale hydrothermal alteration, and the distribution and origin of Cordilleran Fe-oxide(-Cu-Au-P-REE) systems. *XIV Congreso Geológico Chileno*.
- Barton, M. D. (2013). Iron oxide (-Cu-Au-REE-P-Ag-U-Co) systems. In *Geochemistry of mineral deposits* (pp. 515-541). Elsevier Inc.
- Barton, M. D., & Johnson, D. A., (1996). Evaporitic-source model for igneous-related Fe oxide–(REE-Cu-Au-U) mineralization. *Geology*, 24(3), 259-262.

- Barton, M. D., & Johnson, D. A., (2000). Alternative Brine Sources for IOCG systems.
- Barton, M. D., & Johnson, D. A., (2004). Footprints of Fe-oxide (-Cu-Au) systems. SEG 2004. Predictive Mineral Discovery Under Cover. Centre for Global Metallogeny, Spec. Pub. 33, The University of Western Australia, pp. 112–116.
- Barton, M. D., Johnson, D. A., & Porter, T. M. (2000). Alternative brine sources for Fe-oxide (-Cu-Au) systems: Implications for hydrothermal alteration and metals. Hydrothermal iron oxide copper-gold and related deposits: A global perspective, 1, 43-60.
- Bashir, D., Montañez, G. D., Sehra, S., Segura, P. S., & Lauw, J. (2020). An Information-Theoretic Perspective on Overfitting and Underfitting. In Australasian Joint Conference on Artificial Intelligence (pp. 347-358). Springer, Cham.
- Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD explorations newsletter, 6(1), 20-29.
- Battles, D. A., & Barton, M. D. (1995). Arc-related sodic hydrothermal alteration in the western United States. *Geology*, 23(10), 913-916.
- Bau, M. (1999). Scavenging of dissolved yttrium and rare earths by precipitating iron oxyhydroxide: experimental evidence for Ce oxidation, Y-Ho fractionation, and lanthanide tetrad effect. *Geochimica et Cosmochimica Acta*, 63(1), 67-77.
- Bau, M., & Dulski, P. (1996). Distribution of yttrium and rare-earth elements in the Penge and Kuruman iron-formations, Transvaal Supergroup, South Africa. *Precambrian Research*, 79(1-2), 37-55.
- Beaudry B. J. & Geschneider K. A. Jr. (1978). Preparation and basic properties of the rare earth metals. In: Gschneidner KA Jr, Eyring L (eds) *Handbook on the physics and chemistry of rare earths*, chapter 2. North-Holland Publishing Company, pp. 174–232.
- Bell, C. M. (1982). The Lower Paleozoic metasedimentary basement of the Coastal Ranges of Chile between 25°30' and 27° S. *Revista geológica de Chile: An international journal on andean geology*, (17), 21-24.

- Bell, C. M. (1984). Deformation produced by the subduction of a Palaeozoic turbidite sequence in northern Chile. *Journal of the Geological Society*, 141(2), 339-347.
- Bellman, R. E. (1978). *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company.
- Benavides, J. (2007). Iron oxide-copper-gold deposits in the Mantoverde area, northern Chile: Ore genesis and exploration guidelines. Tesis de doctorado en geología, Departamento de Ciencias Geológicas e Ingeniería Geológica, Universidad de Queensland: 1-360.
- Benavides, J., Kyser, T. K., Clark, A. H., Oates, C. J., Zamora, R., Tarnovschi, R., & Castillo, B. (2007). The Mantoverde iron oxide-copper-gold district, III Región, Chile: the role of regionally derived, nonmagmatic fluids in chalcopyrite mineralization. *Economic Geology*, 102(3), 415-440.
- Benko, A., & Lányi, C. S. (2009). History of artificial intelligence. In *Encyclopedia of Information bScience and Technology*, Second Edition (pp. 1759 - 1762). IGI Global.
- Berg, K., & Baumann, A. (1985). Plutonic and metasedimentary rocks from the Coastal Range of northern Chile: RbSr and UPb isotopic systematics. *Earth and Planetary Science Letters*, 75(2-3), 101-115.
- Berger, A., Janots, E., Gnos, E., Frei, R., & Bernier, F. (2014). Rare earth element mineralogy and geochemistry in a laterite profile from Madagascar. *Applied geochemistry*, 41, 218-228.
- Bertsekas, D. P., & Tsitsiklis, J. N. (1996). *Neuro-dynamic programming*. Athena Scientific.
- Berzelius J. J., (1824). Undersökning af några Mineralier. 1. Phosphorsyrad Ytterjord. *Kongliga Svenska Vetenskaps-Akademiens Handlingar* 2:334–338.
- Berzelius J. J., (1825). Account of two newly discovered mineral species. *Edinb J Sci* 3:327–332.
- Bezdek, J. C., Coray, C., Gunderson, R., & Watson, J. (1981). Detection and characterization of cluster substructure i. linear structure: Fuzzy c-lines. *SIAM Journal on Applied Mathematics*, 40(2), 339-357.
- Bhavsar, H., & Ganatra, A. (2012). A comparative study of training algorithms for supervised machine learning. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(4), 2231-2307.

- Bilenker, L., Simon, A., Reich, M., Lundstrom, C., Gajos, N., Bindeman, I., Barra, F., Munizaga, R. (2016). Fe-O stable isotope pairs elucidate a high-temperature origin of Chilean iron oxide-apatite deposits. *Geochimica et Cosmochimica Acta* 177, pp. 94-104.
- Bindeman, I. N., & Bailey, J. C. (1999). Trace elements in anorthite megacrysts from the Kurile Island Arc: a window to across-arc geochemical variations in magma compositions. *Earth and Planetary Science Letters*, 169(3-4), 209-226.
- Bishop, C. M. (2006). Pattern recognition. *Machine learning*, 128(9).
- BNamericas. (2020, 1 octubre). BNamericas – Chile impulsara minería de tierras raras. BNamericas. Recuperado 2 de febrero de 2022, de <https://www.bnamericas.com/es/analisis/chile-impulsara-mineria-de-tierras-raras>
- Boden, M. A. (2017). *Inteligencia artificial*. Turner
- Bonson, C. G. (1998). *Fracturing, fluid processes and mineralisation in the Cretaceous continental magmatic arc of Northern Chile (25° 15'-27° 15'S)* (Doctoral dissertation, Kingston University).
- Bonyadi, Z., Davidson, G. J., Mehrabi, B., Meffre, S., & Ghazban, F. (2011). Significance of apatite REE depletion and monazite inclusions in the brecciated Se–Chahun iron oxide–apatite deposit, Bafq district, Iran: insights from paragenesis and geochemistry. *Chemical Geology*, 281(3-4), 253-269.
- Bookstrom, A. A. (1977). The magnetite deposits of El Romeral, Chile. *Economic Geology*, 72(6), 1101-1130.
- Boric, R., Díaz, F., & Maksaev, V. (1990). *Geología y yacimientos metalíferos de la region de Antofagasta*. SERNAGEOMIN Bol 40:1-246.
- Bournas, N., Touré, A., Balboné, M., Zagré, P. S., Ouédraogo, A., Khaled, K., Prikhodiko, A., & Legault, J. (2019). Use of machine learning techniques on airborne geophysical data for mineral resources exploration in Burkina Faso. *ASEG Extended Abstracts*, 2019(1), 1-4.
- Bousquet, O., Boucheron, S. & Lugosi, G. (2004). 'Introduction to Statistical Learning Theory', in O. Bousquet, U. von Luxburg & G. Ratsch (eds), *Advanced Lectures on Machine*

- Learning Lecture Notes in Artificial Intelligence 3176, Springer, Heidelberg, Germany, pp. 169-207.
- Brachman, R. J., & Anand, T. (1996). The process of knowledge discovery in databases. In *Advances in knowledge discovery and data mining* (pp. 37-57).
- Bramer, M. (2007). *Principles of data mining* (Vol. 180). London: Springer.
- Braverman, D. (1962). Learning filters for optimum pattern recognition. *IRE Transactions on Information Theory*, 8(4), 280-285.
- Bravo, G. J. (2019). *Pymach y Sparkmach: sistemas de procesamiento de datos con dimensión variable usando algoritmos de Machine Learning*. [Tesis Profesional, Universidad Nacional de Ingeniería]. <http://cybertesis.uni.edu.pe/handle/uni/18949>
- Breck, E., Polyzotis, N., Roy, S., Whang, S., & Zinkevich, M. (2019). Data Validation for Machine Learning. In *MLSys*.
- Bressert, E. (2012). *SciPy and NumPy: an overview for developers*. O'Reilly & Associates, Sebastopol, CA.
- Brown, M., Diaz, F., & Grocott, J. (1993). Displacement history of the Atacama fault system 25° 00' S-27° 00' S, northern Chile. *Geological Society of America Bulletin*, 105(9), 1165-1174.
- Bruening, P. J., & Waterman, K. K. (2010). Data tagging for new information governance models. *IEEE Security & Privacy*, 8(5), 64-68.
- Burl, M. C., Asker, L., Smyth, P., Fayyad, U., Perona, P., Crumpler, L., & Aubele, J. (1998). Learning to recognize volcanoes on Venus. *Machine Learning*, 30(2), 165-194.
- Burnham, C.W., (1979). Magmas and hydrothermal fluids, in Barnes, H.L., ed., *Geochemistry of hydrothermal ore deposits*, 2nd edition: New York, John Wiley and Sons, p. 71–136.
- Butler, C. J. (2014). Rare earth elements: China's monopoly and implications for US national security. *Fletcher F. World Aff.*, 38, 23.
- Campesato, O. (2020). *Artificial intelligence, machine learning, and deep learning*. Mercury Learning and Information.

- Caracheo, F. (2002). Modelo educativo (propuesta de diseño). Dirección General de Institutos Tecnológicos. Coordinación Sectorial de Normatividad Académica. México: CIDET.
- Cartagena, C (2021, 30 noviembre). cluster_baseline. Kaggle.com. Recuperado el 02 de noviembre de 2022 de <https://www.kaggle.com/code/cristianminas/cluster-baseline>
- Carten, R. B. (1981). Sodium-calcium metasomatism and its time-space relationship to potassium metasomatism in the Yerington porphyry copper deposit, Lyon County, Nevada. Stanford University.
- Castor, S. B., & Hedrick, J. B. (2006). Rare earth elements. Industrial minerals and rocks, 769-792.
- Cembrano, J., González, G., Arancibia, G., Ahumada, I., Olivares, V., & Herrera, V. (2005). Fault zone development and strain partitioning in an extensional strike-slip duplex: A case study from the Mesozoic Atacama fault system, Northern Chile. Tectonophysics 400, pp. 105-125.
- Chakhmouradian, A. R., & Zaitsev, A. N. (2012). Rare earth mineralization in igneous rocks: sources and processes. Elements, 8(5), 347-353.
- Chan, K. Y., Kwong, C. K., Dillon, T. S., & Tsim, Y. C. (2011). Reducing overfitting in manufacturing process modeling using a backward elimination based genetic programming. Applied Soft Computing, 11(2), 1648-1656.
- Chapelle, O., Schölkopf, B., & Zien, A. (2006). Semi-supervised learning (1st ed.). Cambridge: The MIT Press.
- Charniak, E. & McDermott, D. (1985). Introduction to Artificial Intelligence. Addison-Wesley.
- Charrier, R., Pinto, L., & Rodríguez, M.P. (2007). Tectonostratigraphic evolution of the Andean orogen in Chile: the geology of Chile. The Geological Society of London, London, pp. 21–114.
- Chávez, T. (2017). Estudio de la factibilidad técnica y económica para el diseño de una planta de extracción de minerales estratégicos a partir de pasivos ambientales mineros. [Título profesional, Universidad Técnica Federico Santa María]. Repositorio Académico

- Chen, W. T., & Zhou, M. F. (2014). Ages and compositions of primary and secondary allanite from the Lala Fe–Cu deposit, SW China: implications for multiple episodes of hydrothermal events. *Contributions to Mineralogy and Petrology*, 168, 1-20.
- Chen, W. T., & Zhou, M. F. (2015). Mineralogical and geochemical constraints on mobilization and mineralization of rare Earth elements in the Lala Fe-Cu-(Mo, Ree) deposit, SW China. *American Journal of Science*, 315(7), 671-711.
- Cheung, Y. M. (2003). k*-Means: A new generalized k-means clustering algorithm. *Pattern Recognition Letters*, 24(15), 2883-2893. Classification. Department of Computer Science, National Taiwan University,
- Chiaradia, M., Banks, D., Cliff, R., Marschik, R., & De Haller, A. (2006). Origin of fluids in iron oxide–copper–gold deposits: constraints from $\delta^{37}\text{Cl}$, $^{87}\text{Sr}/^{86}\text{Sr}$ and Cl/Br. *Mineralium Deposita*, 41(6), 565-573.
- Childress, T. M., Simon, A. C., Reich, M., Barra, F., Arce, M., Lundstrom, C. C., & Bindeman, I. N. (2020). Formation of the Mantoverde iron oxide-copper-gold (IOCG) deposit, Chile: insights from Fe and O stable isotopes and comparisons with iron oxide-apatite (IOA) deposits. *Mineralium Deposita*, 55(7), 1489-1504.
- Choi, J. (2018). A study on the standardization strategy for building of learning data set for machine learning applications. *Journal of Digital Convergence*, 16(10), 205-212.
- Chou, I. M., & Eugster, H. P. (1977). Solubility of magnetite in supercritical chloride solutions. *American Journal of Science*, 277(10), 1296-1314.
- Cisternas, M. E. (1986). Stratigraphische, fazielle und lithogeochemische Untersuchungen in der Unterkreide der Region Atacama: metallogenetische Bedeutung am Beispiel der schichtgebundenen Eisen-Lagerstätte Bandurrias. *Heidelberger Geowissenschaftliche Abhandlungen* 2, 268 p.
- Cisternas, M. E., & Díaz, L. L. (1990). Geologic evolution of the Atacama Basin during the Lower Cretaceous. In *Stratabound ore deposits in the Andes* (pp. 495-504). Springer, Berlin, Heidelberg.

- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.
- Coira, B., Davidson, C., Mpodozis, C., Ramos, V. (1982). Tectonic and magmatic evolution of the Andes of northern Argentina and Chile. *Earth-Science Reviews*, Special Issue, 18, pp. 303–332.
- Cole, C. S., James, R. H., Connelly, D. P., & Hathorne, E. C. (2014). Rare earth elements as indicators of hydrothermal processes within the East Scotia subduction zone system. *Geochimica et Cosmochimica Acta*, 140, 20-38.
- Comisión Chilena del Cobre, COCHILCO. (2016). Situación actual del mercado de tierras raras y su potencial en Chile. Santiago, Chile.
- CORFO. (2013). Concurso bienes públicos para la competitividad 2013: “Identificar elementos de valor en residuos mineros (relaves) y evaluar su recuperación como productos comerciales”. pp. 228.
- Cornejo, P., Matthews, S., Orrego, M., & Robles, W. (2000). Etapas de mineralización asociadas a alteración potásica en un sistema Fe-Cu-Au: Yacimiento Mantoverde, III Región de Atacama, Chile: IX Congreso Geológico Chileno, Puerto Varas. *Actas*, 97-101.
- Cortés, J. I. M., & del Valle Giles, A. V. (2014). Las tierras raras: un sector estratégico para el desarrollo tecnológico de China. [Título profesional Universidad Nacional Autónoma de México]. Repositorios Académico Universidad Nacional Autónoma de México. https://ru.dgb.unam.mx/handle/DGB_UNAM/TES01000716809
- Corvalán, J. (1974). Estratigrafía del Neocomiano marino de la región al sur de Copiapó, Provincia de Atacama. *Revista Geológica de Chile* 1, 13-36.
- Costa, I. S. L., Tavares, F. M., & de Oliveira, J. K. M. (2019). Predictive lithological mapping through machine learning methods: a case study in the Cinzento Lineament, Carajás Province, Brazil. *Journal of the Geological Survey of Brazil*, 2(1), 26-36.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.
- Cracknell, M. J. (2014). Machine learning for geological mapping: Algorithms and applications (Doctoral dissertation, University of Tasmania).

- Craddock, P. R., Bach, W., Seewald, J. S., Rouxel, O. J., Reeves, E., & Tivey, M. K. (2010). Rare earth element abundances in hydrothermal fluids from the Manus Basin, Papua New Guinea: Indicators of sub-seafloor hydrothermal processes in back-arc basins. *Geochimica et Cosmochimica Acta*, 74(19), 5494-5513.
- Creixell, C., Arévalo, C., & Fanning, M. (2009). Geochronology of the cretaceous magmatism from the Coastal Cordillera of north-central Chile (29°15' to 29°30'S): metallogenic implications. XII Congreso Geológico Chileno.
- Periodni (s.f.) Download *tabla_periodica-4k-3840x2160-fondo_oscuro.png* image from www.periodni.com. Copyright © 1998-2023 Eni Generalic, All Rights Reserved. https://www.periodni.com/es/download_image.php?name=tabla_periodica-4k-3840x2160-fondo_oscuro.png&source=es/images.html
- Dallmeyer, R. D., Brown, M., Grocott, J., Taylor, G. K., & Treloar, P. J. (1996). Mesozoic magmatic and tectonic events within the Andean plate boundary zone, 26-27 30'S, north Chile: Constraints from ⁴⁰Ar/³⁹Ar mineral ages. *The Journal of Geology*, 104(1), 19-40.
- Dalziel, I.W.D., Grunow, A.M., Storey, B.C., Garrett, S.W., Herrod, L.D.B., Pankhurst, R.J. (1987). Extensional Tectonics and the Fragmentation of Gondwanaland. Geological Society of London, pp. 433–441.
- Dare, S., Barnes, S.-J., Beaudin, G., (2014). Did the massive magnetite “lava flows” of El Laco (Chile) form by magmatic or hydrothermal processes? New constraints from magnetite composition by LA-ICP-MS. *Min Dep* 50, 607–617. <https://doi.org/10.1007/s00126-014-0560-1>.
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), 224-227.
- Daviran, M., Maghsoudi, A., Ghezelbash, R., & Pradhan, B. (2021). A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach. *Computers and Geosciences*, 148. <https://doi.org/10.1016/j.cageo.2021.104688>
- Davis, D. (2021, junio). Machine Learning in Python – The Ttop New Scikit-Learn 0.24 Features You Should Know. freeCodeCamp.Org. Recuperado 25 de Agosto de 2022, de

<https://www.freecodecamp.org/news/machine-learning-python-new-scikit-learn-features-you-should-know/>

- de Haller, A. & Fontboté, L. (2009). The rauc-condestable iron oxide copper-gold deposit, central coast of peru: Ore and related hydrothermal alteration, sulfur isotopes, and thermodynamic constraints. *Econ. Geol.* 104, 365–384.
- de Haller, A., Zúñiga, A. J., Corfu, F., & Fontboté, L. (2002). The iron oxide-Cu-Au deposit of Raúl-Condestable. In Mala, Lima, Peru [abs.]: Congreso Geológico Peruano, 11th, Resúmenes (p. 80).
- de la Hoz, G. M. (2012). Las tierras raras: elementos claves del siglo XXI. *Revista cuatrimestral de del instituto de Bio y Geociencias*. Año II, N°2.
- Deer W. A., Howie R. A., & Zussman J., (1986a). Eudialyte-Eucolite. The rockforming minerals, volumen 1B, disilicates and ringsilicates, 2nd edn. Longman Scientific and Technical, Harlow, pp 348–363.
- Deer W. A., Howie R. A., & Zussman J., (1986b). Allanite. The rockforming minerals, volume 1B, disilicates and ringsilicates, 2nd edn. The Geological Society, Bath, United Kingdom, pp 151–179.
- del Real, I., Reich, M., Simon, A. C., Deditius, A., Barra, F., Rodríguez-Mustafa, M. A., Thompson, J & Roberts, M. P. (2021). Formation of giant iron oxide-copper-gold deposits by superimposed, episodic hydrothermal pulses. *Communications Earth & Environment*, 2(1), 1-9.
- del Real, I., Thompson, J. F. H., & Carriedo, J. (2018). Lithological and structural controls on the genesis of the Candelaria-Punta del Cobre Iron Oxide Copper Gold district, Northern Chile. *Ore Geology Reviews*, 102, 106–153. <https://doi.org/10.1016/j.oregeorev.2018.08.034>
- del Real, I., Thompson, J. F. H., Simon, A. C., & Reich, M. (2020). Geochemical and isotopic signature of pyrite as a proxy for fluid source and evolution in the Candelaria-Punta del Cobre iron oxide copper-gold district, Chile. *Economic Geology*, 115(7), 1493-1518.
- Dietterich, T. G., & Kong, E. B. (1995). Machine learning bias, statistical bias, and statistical variance of decision tree algorithms (pp. 0-13). Technical report, Department of Computer Science, Oregon State University.

- Dmitrijeva, M., Ciobanu, C. L., Ehrig, K. J., Cook, N. J., Metcalfe, A. v., Verdugo-Ihl, M. R., & McPhie, J. (2019b). Mineralization-alteration footprints in the Olympic Dam IOCG district, South Australia: The Acropolis prospect. *Journal of Geochemical Exploration*, 205. <https://doi.org/10.1016/j.gexplo.2019.106333>
- Dmitrijeva, M., Ehrig, K., Ciobanu, C., Cook, N., Verdugo-Ihl, M., Metcalfe, A., (2019a). Defining IOCG signatures through compositional data analysis: a case study of lithogeochemical zoning from the Olympic Dam deposit, South Australia. *Ore Geol. Rev.* 105, 86–101. <https://doi.org/10.1016/j.oregeorev.2018.12.013>.
- Domingos, P. (2000). Bayesian averaging of classifiers and the overfitting problem. In *ICML* (Vol. 747, pp. 223-230).
- Duan, C., Li, Y., Mao, J., Wan, Q., He, S., Wang, C., Yang, B., & Hou, K. (2023). Zircon U–Pb ages, Hf–O isotopes and trace elements of the multi-volcanism in the Ningwu ore district, eastern China: Implications for the magma evolution and fertility of iron oxide–apatite (IOA) deposits. *Gondwana Research*.
- Dutta, T., Kim, K. H., Uchimiya, M., Kwon, E. E., Jeon, B. H., Deep, A., & Yun, S. T. (2016). Global demand for rare earth resources and strategies for green mining. *Environmental Research*, 150, 182-190.
- Ehrig, K., McPhie, J., Kamenetsky, V., (2012). Geology and mineralogical zonation of the Olympic Dam iron oxide Cu-U-Au-Ag deposit, South Australia. *Soc. Econ. Geol. Spec. Pub.* 16, 237–267. <https://doi.org/10.5382/SP.16.11>.
- Ehya, F. (2012). Variation of mineralizing fluids and fractionation of REE during the emplacement of the vein-type fluorite deposit at Bozijan, Markazi Province, Iran. *Journal of Geochemical Exploration*, 112, 93-106.
- El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? In *Machine Learning in Radiation Oncology* (pp. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3_1.
- Emsbo, P., McLaughlin, P. I., Breit, G. N., du Bray, E. A., & Koenig, A. E. (2015). Rare earth elements in sedimentary phosphate deposits: Solution to the global REE crisis?. *Gondwana Research*, 27(2), 776-785.

- Espinoza, S. (1985). Distribución de elementos de transición en magnetitas de yacimientos de hierro Chilenos y su significado en la tipología y génesis probable de los yacimientos. In Congreso Geológico Chileno, 4th, Antofagasta, Actas (p. 3).
- Espinoza, S. (1990). The Atacama-Coquimbo ferriferous belt, northern Chile. In *Stratabound ore deposits in the Andes* (pp. 353-364). Springer, Berlin, Heidelberg.
- Espinoza, S., Veliz, H., Esquivel, J., Arias, J., & Moraga, A. (1996). The cupriferous province of the Coastal Range, northern Chile. Petersen R (eds) *Andean copper deposits: New discoveries, mineralization styles and metallogeny*. Soc Econ Geol Spec Publ 5:19–32
- Evans, D., & Eysers, D. M. (2012). Efficient data tagging for managing privacy in the internet of things. In 2012 IEEE International Conference on Green Computing and Communications (pp. 244-248). IEEE.
- Fahim, A. M., Salem, A. M., Torkey, F. A., & Ramadan, M. (2006). An efficient enhanced k-means clustering algorithm. *Journal of Zhejiang University-Science A*, 7(10), 1626-1633.
- Fan, H. R., Yang, K. F., Hu, F. F., Liu, S., & Wang, K. Y. (2016). The giant Bayan Obo REE-Nb-Fe deposit, China: controversy and ore genesis. *Geoscience Frontiers*, 7(3), 335-344.
- Fayyad, U. M, Piatetsky-Shapiro, G., & Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
- Fayyad, U. M, Piatetsky-Shapiro, G., & Smyth, P. (1996b). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In *KDD* (Vol. 96, pp. 82-88).
- Fayyad, U. M, Piatetsky-Shapiro, G., & Smyth, P. (1996c). From Data Mining to Knowledge Discovery: An Overview. *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1-34.
- Fayyad, U. M. & Simoudis, E. (1997). Data Mining and Knowledge Discovery in Databases: Implications. In *Ninth International Conference on Scientific and Statistical* (pp. 2-11). IEEE Computer Society.
- Feizollah, A., Anuar, N. B., Salleh, R., & Amalina, F. (2014). Comparative study of k-means and mini batch k-means clustering algorithms in android malware detection using network

- traffic analysis. In 2014 international symposium on biometrics and security technologies (ISBAST) (pp. 193-197). IEEE.
- Felsche, J., (1971). The crystal structures of the dimorphic rare earth disilicate, $\text{Pr}_2\text{Si}_2\text{O}_7$. *Zeitschrift für Kristallographie-Crystalline Materials*, 133(1-6), 364-385.
- Felsche, J., (1973). The crystal chemistry of the rare-earth silicates. In *Rare earths* (pp. 99-197). Springer, Berlin, Heidelberg.
- Felsche, J., (1978). Yttrium and lanthanides. Crystal chemistry. In: K.H. Wedepohl (Editor), *Handbook of Geochemistry*, 11/5. Springer-Verlag, Berlin, pp. 39,57-71-A-1 to 39,57-71A-42.
- Feng, C. & Michie, D. (1994). *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, New York, pp. 50-83.
- Foley, N. K., & Ayuso, R. A. (2013). Rare earth element mobility in high-alumina altered metavolcanic deposits, South Carolina, USA. *Journal of Geochemical Exploration*, 133, 50-67.
- Fradkov, A. L. (2020). Early history of machine learning. *IFAC-PapersOnLine*, 53(2), 1385-1390.
- Franchini, M., Lira, R., Meinert, L., Ríos, F. J., Poklepovic, M. F., Impiccini, A., & Millone, H. A. (2005). Na-Fe-Ca alteration and LREE (Th-Nb) mineralization in marble and granitoids of Sierra de Sumampa, Santiago del Estero, Argentina. *Economic Geology*, 100(4), 733-764.
- Franklin, J.M., Gibson, H.L., Jonasson, I.R., & Galley, A.G., (2005). Volcanogenic massive sulfide deposits: *ECONOMIC GEOLOGY 100TH ANNIVERSARY VOLUME*, p. 523–560.
- Fredes, M. J. (2017). Geología del prospecto IOCG Alcaparra D, comuna de Huasco, Región de Atacama. [Título Profesional, Universidad de Chile]. Repositorio Académico de la Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/148184>
- French, R. M. (2000). The Turing Test: the first 50 years. *Trends in cognitive sciences*, 4(3), 115-122.

- Frietsch, R., (1978). On the magmatic origin of iron ores of the Kiruna type. *Econ. Geol.* 73, 478–485. <https://doi.org/10.2113/gsecongeo.73.4.478>.
- Frietsch, R., (1979). The occurrence and composition of apatite with special reference to iron ores and rocks in northern Sweden. *Sver. Geol. Unders.*, 694C: 49 pp.
- Gabás, N. E. G. (2020). Consecuencias geoestratégicas de la hegemonía china en el mercado de las tierras raras. *Global strategy reports*, (43), 1. <https://global-strategy.org/consecuencias-geoestrategicas-de-la-hegemonia-china-en-el-mercado-de-las-tierras-raras/>
- García, F. (1967). Geología del norte grande de Chile. In *Symp. sobre el geosinclinal andino* (Vol. 3, p. 138). Soc. Geol. Chile.
- García-Bilbao P. (2015). Las tierras raras: encrucijada de conflictos.
- Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Springer.
- Gass, S., & Saaty, T. (1955). The computational algorithm for the parametric objective function. *Naval research logistics quarterly*, 2(1-2), 39-45.
- Gazley, M. F., Collins, K. S., Roberston, J., Hines, B. R., Fisher, L. A., & McFarlane, A. (2015). Application of principal component analysis and cluster analysis to mineral exploration and mine geology. In *AusIMM New Zealand branch annual conference* (Vol. 2015, pp. 131-139). Dunedin New Zealand.
- Geijer, P., (1931). The iron ores of the Kiruna type: geographical distribution, geological characters, and origin. *Sver. Geol. Unders.*, Ser 39.
- Gelcich, S., Davis, D.W., & Spooner, E.T.C. (2005). Testing the apatitemagnetite geochronometer: U-Pb and $^{40}\text{Ar}/^{39}\text{Ar}$ geochronology of plutonic rocks, massive magnetite-apatite tabular bodies, and IOCG mineralization in northern Chile: *Geochimica et Cosmochimica Acta*, v. 69, p. 3367–3384.
- Gelfort, R. (2006). On classification of logging data (Doctoral dissertation, Master's thesis, Clausthal University of Technology, Germany).
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc."

- Gipson, M. (2002). Palaeomagnetism and geochronology of the Andean forearc in the Vallenar district (298S), Region III, northern Chile. PhD thesis, University of Plymouth.
- Glesson, D. (2019, 28 junio). OZ Minerals Explorer Challenge winners crowned. International Mining. Recuperado 26 de abril de 2022, de <https://im-mining.com/2019/06/28/oz-minerals-explorer-challenge-winners-crowned/>
- Godoy, E., Lara, L., & Ugalde, H. (1997). La falla Chivato: borde oriental del plutonismo asociado al sistema de falla Atacama. In VII Congreso Geológico Chileno, Universidad Católica del Norte, Antofagasta, Actas (Vol. 3, pp. 1604-1607).
- Goldfarb, R.J., Groves, D.I., & Gardoll, S., (2001). Orogenic gold and geological time: A global synthesis: Ore Geology Reviews, v. 18, p. 1–75.
- Goleman, A. Daniel; Boyatzis, Richard; Mckee, (2019). Hands-on Machine Learning, vol. 53, no. 9.
- Gonzales, L. (2018a, 9 noviembre). Introducción a Bias y Varianza. AprendeIA. Recuperado 26 de octubre de 2021, de <https://aprendeia.com/bias-y-varianza-en-machine-learning/>.
- Gonzales, L. (2018b, 16 noviembre). Introducción a Sobreajuste y Subajuste para Machine Learning. AprendeIA. Recuperado 25 de noviembre de 2021, de <https://aprendeia.com/sobreajuste-y-subajuste-en-machine-learning/>.
- Goodenough, K. M. (2017). Verplanck and Hitzman: Rare earth and critical elements in ore deposits.
- Goonan, T. G. (2011). Rare earth elements: End use and recyclability (p. 19Disponible). Reston, VA, USA: US Department of the Interior, US Geological Survey.
- Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press.
- Gow, P.A., Wall, V.J., Oliver, N.H.S., Valenta, R.K., (1994). Proterozoic iron oxide (Cu-U-Au-REE) deposits: Further evidence of hydrothermal origins. Geology, vol. 22, p. 633-636.
- Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. Journal of classification, 3, 5-48.
- Granath, G. (1988). Pattern recognition in geochemical hydrocarbon exploration: a fuzzy approach. Mathematical geology, 20, 673-691.

- Gratz, R., & Heinrich, W. (1997). Monazite-xenotime thermobarometry: Experimental calibration of the miscibility gap in the binary system CePO₄-YPO₄. *American Mineralogist*, 82(7-8), 772-780.
- Gray.D. (2019). “Introducing Voyage Deepdrive - Voyage.” [Online]. Available: <https://news.voyage.auto/introducing-voyage-deepdrive-69b3cf0f0be6>. [Accessed: 20-Oct-2021].
- Grebby, S., Naden, J., Cunningham, D., & Tansey, K. (2011). Integrating airborne multispectral imagery and airborne LiDAR data for enhanced lithological mapping in vegetated terrain. *Remote Sensing of Environment*, 115(1), 214-226.
- Gregory, D. D., Large, R. R., Halpin, J. A., Baturina, E. L., Lyons, T. W., Wu, S., ... & Bull, S. W. (2015). Trace element content of sedimentary pyrite in black shales. *Economic Geology*, 110(6), 1389-1410.
- Grocott, J., & Taylor, G. (2002). Magmatic arc fault systems, deformation partitioning and emplacement of granitic complexes in the Coastal Cordillera, north Chilean Andes (25°30'S to 27°00'S). *Journal of the Geological Society, London* 159: 425-442.
- Grocott, J., Brown, M., Dallmeyer, R.D., Taylor, G.K., & Treloar, P.J., (1994). Mechanisms of continental growth in extensional arcs: an example from the Andean plate-boundary zone. *Geology* 22, 391–394.
- Groves, D.I., & Bierlein, F.P., (2007). Geodynamic settings of mineral deposit systems: *Journal of the Geological Society, London*, v. 164, p. 19–30.
- Groves, D.I., Bierlein, F.P., Meinert, L.D., Hitzman, M.W., (2010). Iron oxide copper-gold (IOCG) deposits through Earth history: implications for origin, lithospheric setting, and distinction from other epigenetic iron oxide deposits. *Econ. Geol.* 105 (3), 641–654. <https://doi.org/10.2113/gsecongeo.105.3.641>.
- Groves, D.I., Condie, K.C., Goldfarb, R.J., Hronsky, J.M.A., & Vielreicher, R.M., (2005a). Secular changes in global tectonic processes and their influence on the temporal distribution of gold-bearing mineral deposits: *ECONOMIC GEOLOGY*, v. 100, p. 203–224.

- Groves, D.I., Mair, J.L., Vielreicher, R.M., & Hart, C.J.R., (2005b). The lithospheric settings of Carlin-type deposits: An important clue to their genetic associations and deposit associates: Geological Society of Nevada Symposium 2005, Window on the World, Reno, Nevada, Proceedings, p. 1–11.
- Groves, D.I., Vielreicher, R.M., Goldfarb, R.J. & Condie, K.C., (2005c). Controls on the heterogeneous distribution of mineral deposits through time, in Mc Donald, I. et al., eds., Mineral deposits and earth evolution: Geological Society, London, Special Publication, v. 248, p. 71–101.
- Gubbins, D. (2004). Time Series Analysis and Inverse Theory for Geophysicists, Cambridge University Press, Cambridge, p. 255.
- Gupta C. K. & Krishnamurthy N. (2005). Extractive metallurgy of rare earths. CRC Press. 484 p. International Materials Reviews, 37:1, 197-248, [DOI: 10.1179/imr.1992.37.1.197](https://doi.org/10.1179/imr.1992.37.1.197)
- Guyon, X., & Yao, J. F. (1999). On the underfitting and overfitting sets of models chosen by order selection criteria. Journal of Multivariate Analysis, 70(2), 221-249.
- Gyorfi, L., Ottucsak, G., & Walk, H. (Eds.). (2012). Machine learning for financial engineering (Vol. 8). World Scientific.
- Hacking, I. (1983). Representing and intervening: Introductory topics in the philosophy of natural science. Cambridge university press.
- Hai-qing, Y. A. N., Bao-lin, H. E., Qiao-feng, L. I. U., Hui-min, J. I. A., & Fei, L. I. (2014). Characteristics of Magmatic Geochronology and Geochemistry of Dashui Gold Deposit in West Qinling. Journal of Earth Sciences & Environment, 36(1).
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014). Pearson new international edition. Multivariate data analysis, Seventh Edition. Pearson Education Limited Harlow, Essex.
- Hajnajafi, G., Jafarirad, A., Afzal, P., & Sheikh-Zakariaee, S. J. (2021). Geological interpretation using multivariate K-means and robust factor analysis in Dezak area, SW Iran. Environmental Earth Sciences, 80(1), 1-13.
- Han, J., Kamber, M., & Pei, J., (2012). Data Mining, Southeast Asia Edition: Concepts and Techniques. Morgan kaufmann.

- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hand, M., Reid, A., & Jagodzinski, L., (2007). Tectonic framework and evolution of the Gawler Craton, southern Australia: *ECONOMIC GEOLOGY*, v. 102, p. 1377–1395.
- Hao, J., & Ho, T. K. (2019). Machine learning made easy: a review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics*, 44(3), 348-361.
- Hastie, T., Tibshirani, R. & Friedman, J.H. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd edn, Springer Series in Statistics, Springer, New York, p. 533.
- Hatch, G., & Lifton, J. (2013, 20 mayo). Technological Metals Research. | We Make Your Metal Detecting Easy. Tech Metals Research. Recuperado el 5 de febrero de 2022, de www.techmetalsresearch.com
- Haugeland, J. (Ed.). (1985). *Artificial Intelligence: The Very Idea*. MIT Press
- Haxel G. B., Boore S., & Mayfield S. (2005). U.S. geological survey. Fact Sheet 087-02. Rare earth elements—critical resources for high technology. <http://pubs.usgs.gov/fs/2002/fs087-02/>.
- Haxel, G. (2005). Ultrapotassic mafic dikes and rare earth elements-and barium-rich carbonatite at Mountain Pass, Mojave Desert, Southern California: Summary and field trip localities (Vol. 2005, p. 53). US Geological Survey.
- Haywood, N. (2008). The geodynamic setting and magmatic controls on genesis of world-class IOCG ore deposits. *Geological Society of Australia*, 89, 130.
- Henderson, P. (1984). General geochemical properties and abundances of the rare earth elements. In *Developments in geochemistry* (Vol. 2, pp. 1-32). Elsevier.
- Henderson, P. (1996). The rare earth elements: introduction and review. *Rare earth minerals: chemistry, origin and ore deposits*, 1-19. (1st, ed.). Chapman & Hall.
- Henderson, P. (Ed.). (2013). *Rare earth element geochemistry*. (Vol 2.). Elsevier.
- Herrera, V., Garmendia, P., & Pizarro, R. (2007). *Geología y recursos minerales del Proyecto Diego de Almagro, Región de Atacama*. Informe Interno para Minera Cerro Dominador.

- Hitzman, M.W., (2000). Iron oxide-Cu-Au deposits: what, where, when, and why. In: Porter, T.M. (Ed.), Hydrothermal Iron Oxide Copper-gold and Related Deposits. Australian Mineral Foundation, Adelaide, pp. 9–25.
- Hitzman, M.W., Oreskes, N., Einaudi, M.T., (1992). Geological characteristics and tectonic setting of Proterozoic iron oxide (Cu-U-Au-LREE) deposits. *Precambrian Res.* 58, 241–287.
- Holcomb, S. D., Porter, W. K., Ault, S. V., Mao, G., & Wang, J. (2018). Overview on deepmind and its alphago zero ai. In Proceedings of the 2018 international conference on big data and education (pp. 67-71).
- Hong, S., Zuo, R., Huang, X., & Xiong, Y. (2021). Distinguishing IOCG and IOA deposits via random forest algorithm based on magnetite composition. *Journal of Geochemical Exploration*, 230. <https://doi.org/10.1016/j.gexplo.2021.106859>
- Hopf, S., 1990. The Agustina mine, a volcanic-hosted copper deposit in northern Chile. In: Fontbote´, L., Amstutz, G.C., Cardozo, M., Cedillo, E., Frutos, J. (Eds.), Stratabound ore deposits in the Andes. Special Publication No. 8, Society for Geology Applied to Mineral Deposits, Springer, Berlin, pp. 421–434.
- Hopper, D., Correa, A., & Porter, T. M. (2000). The Panulcillo and Teresa de Colmo copper deposits: Two contrasting examples of Fe-ox Cu-Au mineralization from the coastal Cordillera of Chile. Hydrothermal iron oxide copper-gold and related deposits: A global perspective, 1, 177-189.
- Howden, W. E. (1975). Methodology for the generation of program test data. *IEEE Transactions on computers*, 100(5), 554-560.
- Huang, X. W., Boutroy, É., Makvandi, S., Beaudoin, G., Corriveau, L., & De Toni, A. F. (2019). Trace element composition of iron oxides from IOCG and IOA deposits: relationship to hydrothermal alteration and deposit subtypes. *Mineralium Deposita*, 54(4), 525-552.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2, 193-218.
- Humphries, M. (2011). Rare-earth elements: The global supply chain: Library of Congress. In Congressional Research Service (CRS) Report for Congress (Vol. 41347, p. 26).

- Huo, H., Rong, Z., Kononova, O., Sun, W., Botari, T., He, T., Tshitoyan, V., & Ceder, G. (2019). Semi-supervised machine-learning classification of materials synthesis procedures. *npj Computational Materials*, 5(1), 1-7.
- Hurst, C. (2010). China's rare earth elements industry: What can the west learn? . Institute for the Analysis of Global Security Washington DC.
- Hurtig, N. C., & Williams-Jones, A. E. (2014). An experimental study of the transport of gold through hydration of AuCl in aqueous vapour and vapour-like fluids. *Geochimica et Cosmochimica Acta*, 127, 305-325.
- Huston, D. L., Sie, S. H., Suter, G. F., Cooke, D. R., & Both, R. A. (1995). Trace elements in sulfide minerals from eastern Australian volcanic-hosted massive sulfide deposits; Part I, Proton microprobe analyses of pyrite, chalcopyrite, and sphalerite, and Part II, Selenium levels in pyrite; comparison with delta 34 S values and implications for the source of sulfur in volcanogenic hydrothermal systems. *Economic Geology*, 90(5), 1167-1196.
- IBM. (2021, 25 marzo). *Underfitting*. Recuperado 10 de marzo de 2022, de <https://www.ibm.com/cloud/learn/underfitting>
- IBM. (s. f.). *Machine Learning*. Recuperado 23 de marzo de 2022, de <https://www.ibm.com/cloud/learn/machine-learning>
- InteractiveChaos. (s. f.). *Conjuntos de datos etiquetados y sin etiquetar | Interactive Chaos*. Recuperado 26 de febrero de 2022, de <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/conjuntos-de-datos-etiquetados-y-sin-etiquetar>
- Ionos. (2019, febrero). Jupyter Notebook: documentos web para análisis de datos, código en vivo y muchos más. IONOS Digitalguide. Recuperado 23 de agosto de 2022, de <https://www.ionos.es/digitalguide/paginas-web/desarrollo-web/jupyter-notebook/>
- Iriarte, S., Arévalo, C., Mpodozis, C., & Rivera, O. (1996). Mapa Geológico de la Hoja Carrera Pinto, escala 1: 1.00. 000. Servicio Nacional de Geología y Minería (Chile), Mapas Geológicos, (3).
- Ismail, R., Ciobanu, C.L., Cook, N.J., Teale, G.S., Giles, D., Schmidt Mumm, A., Wade, B., (2014). Rare earths and other trace elements in minerals from skarn assemblages, Hillside

- iron oxide–copper–gold deposit, Yorke Peninsula, South Australia. *Lithos* 184–187, 456–477. <https://doi.org/10.1016/j.lithos.2013.07.023>.
- Jabbar, H., & Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 163-172.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.
- Jaireth, S., Hoatson, D. M., & Mieзитis, Y. (2014). Geological setting and resources of the major rare-earth-element deposits in Australia. *Ore Geology Reviews*, 62, 72-128.
- Jha, A. R. (2014). *Rare Earth Materials: Properties and Applications*. (1st, ed.). CRC Press.
- Jiyan, C., Ruidong, Y., Huairui, W., & Junbo, G. (2013). Rare earth element geochemistry of Cambrian phosphorites from the Yangtze Region. *Journal of Rare Earths*, 31(1), 101-112.
- Jochum, K.P., Seufert, H.M., Spettel, B. & Palme, H. (1986). The solar-system abundances of Nb, Ta, and Y and the relative abundances of refractory lithophile elements in differentiated planetary bodies. *Geochim. Cosmochim. Acta* 50,1173-1183.
- Jofré, P. A. (2020). Aplicación de técnicas de Machine Learning en el modelamiento de sistemas químicos.
- Jolliffe, I. T. (2002). Principal component analysis for special types of data (pp. 338-372). Springer New York.
- Jones, A. P., Wall, F., & Williams, C. T. (1995). *Rare earth minerals: chemistry, origin and ore deposits* (Vol. 7). Springer Science & Business Media.
- Jorquera, Á. A. (2019). Modelación de la variabilidad espacial de la profundidad de nieve con Machine Learning y Geoestadística utilizando datos.
- Jurgan, H. (1977a). Zur Gliederung der Unterkreide-Serien in der Provinz Atacama Chile. *Geologische Rundschau*, 66(1), 404-434.
- Jurgan, H. (1977b). Strukturelle und lithofazielle entwicklung des andinen Unterkreide-beckens im Norden Chiles (Provinz Atacama): Geotektonische Forschung, v. 52, 138 p.

- Kanazawa, Y., & Kamitani, M. (2006). Rare earth minerals and resources in the world. *Journal of alloys and compounds*, 408, 1339-1343.
- Kashnitsky, K. (2019, 28 julio) Topic 7. Unsupervised learning: PCA and clustering. Kaggle.com. Recuperado el 10 de septiembre de 2022 de [Topic 7. Unsupervised learning: PCA and clustering | Kaggle](#)
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.
- Keith, M., Haase, K. M., Klemm, R., Krumm, S., & Strauss, H. (2016). Systematic variations of trace element and sulfur isotope compositions in pyrite with stratigraphic depth in the Skouriotissa volcanic-hosted massive sulfide deposit, Troodos ophiolite, Cyprus. *Chemical Geology*, 423, 7-18.
- Keith, M., Smith, D. J., Jenkin, G. R., Holwell, D. A., & Dye, M. D. (2017). A review of Te and Se systematics in hydrothermal pyrite from precious metal deposits: Insights into ore-forming processes. *Ore Geology Reviews*, 96, 269-282.
- Kerrick, R., Goldfarb, R.J., & Richards, J.P., (2005). Metallogenic provinces in an evolving dynamic framework: ECONOMIC GEOLOGY 100TH ANNIVERSARY VOLUME, p. 1097–1136.
- Kerrick, R., Goldfarb, R.J., Cline, J., & Leach, D., (2008). Metallogenic provinces of North America in a superplume-supercontinent framework: *Arizona Geological Society Digest*, v. 22, p. 1–18.
- Kerrick, R., Goldfarb, R.J., Groves, D.I., & Garwin, S., (2000). The geodynamics of world-class gold deposits: Characteristics, space-time distribution, and origins: *Reviews in Economic Geology*, v. 13, p. 501–552.
- Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal*, 17(6), 441-458.

- Khanum, M., Mahboob, T., Imtiaz, W., Ghafoor, H. A., & Sehar, R. (2015). A survey on unsupervised machine learning algorithms for automation, classification, and maintenance. *International Journal of Computer Applications*, 119(13).
- Knipping, J. L., Bilenker, L. D., Simon, A. C., Reich, M., Barra, F., Deditius, A. P., Walle, M., Heinrich, C., Holtz, F., & Munizaga, R. (2015a). Trace elements in magnetite from massive iron oxide-apatite deposits indicate a combined formation by igneous and magmatic-hydrothermal processes. *Geochimica et Cosmochimica Acta*, 171, 15-38.
- Knipping, J.L., Bilenker, L.D., Simon, A.C., Reich, M., Barra, F., Deditius, A.P., Lundstrom, C., Bindeman, I., Munizaga, R., (2015b). Giant Kiruna-type deposits form by efficient flotation of magmatic magnetite suspensions. *Geology* 43, 591–594. <https://doi.org/10.1130/G36650.1>.
- Kogarko, L. N., Williams, C. T., & Woolley, A. R. (2006). Compositional evolution and cryptic variation in pyroxenes of the peralkaline Lovozero intrusion, Kola Peninsula, Russia. *Mineralogical Magazine*, 70(4), 347-359.
- Kontonikas-Charos, A., Ciobanu, C. L., Cook, N. J., Ehrig, K., Ismail, R., Krneta, S., & Basak, A. (2018). Feldspar mineralogy and rare-earth element (re) mobilization in iron-oxide copper gold systems from South Australia: a nanoscale study. *ALKIS KONTONIKAS-CHAROS ET AL. FELDSPARS IN IOCG SYSTEMS. Mineralogical Magazine*, 82(S1), S173-S197.
- Kopera, J. J. (2004). Inside the Nickel metal hydride battery. Cobasys, MI, USA http://www.cobasys.com/pdf/tutorial/inside_nimh_battery_technology.pdf (Accessed 08/11).
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.
- Kristanto, N. H., ACL, A., & HB, S. (2016). Implementasi K-Means Clustering untuk Pengelompokan Analisis Rasio Profitabilitas dalam Working Capital. *Jurnal Informatika dan Sistem Informasi*, 2(01), 9-14.

- Krneta, S., Ciobanu, C. L., Cook, N. J., & Ehrig, K. J. (2018). Numerical modeling of REE fractionation patterns in fluorapatite from the Olympic Dam deposit (South Australia). *Minerals*, 8(8), 342.
- Krneta, S., Ciobanu, C.L., Cook, N.J., Ehrig, K., Kontonikas-Charos, A., (2016). Apatite at Olympic Dam, South Australia: a petrogenetic tool. *Lithos* 262, 470–485. <https://doi.org/10.1016/j.lithos.2016.07.033>.
- Krneta, S., Ciobanu, C.L., Cook, N.J., Ehrig, K.J., Kontonikas-Charos, A., (2017a). Rare Earth Element Behaviour in Apatite from the Olympic Dam Cu–U–Au–Ag Deposit, South Australia. *Minerals* 7, 135. <https://doi.org/10.3390/min7080135>.
- Krneta, S., Cook, N.J., Ciobanu, C.L., Ehrig, K.J., Kontonikas-Charos, A., (2017b). The Wirrda Well and Acropolis prospects, Gawler Craton, South Australia: Insights into evolving fluid conditions through apatite chemistry. *J. Geochem. Explor.* 181, 276–291. <https://doi.org/10.1016/j.gexplo.2017.08.004>.
- Krueger, D., Ballas, N., Jastrzebski, S., Arpit, D., Kanwal, M. S., Maharaj, T., Bengio, E, Fischer, A., & Courville, A. (2017). Deep nets don't learn via memorization.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). New York: Springer.
- Kuncheva, L. (2004). *Combining Pattern Classifiers: Methods and Algorithms*, Wiley, Hoboken, New Jersey, p. 376. 'Learning to recognize volcanoes on Venus', *Machine Learning*, vol. 30, no. 2, pp. 165-194.
- Küpelı, Ş. (2010). Trace and rare-earth element behaviors during alteration and mineralization in the Attepe iron deposits (Feke-Adana, southern Turkey). *Journal of Geochemical Exploration*, 105(3), 51-74.
- Kurzweil, R. (1990). *The Age of Intelligent Machines*. MIT Press.
- Kynicky, J., Smith, M. P., & Xu, C. (2012). Diversity of rare earth deposits: the key example of China. *Elements*, 8(5), 361-367.
- Lara, L., & Godoy, E. (1998). *Hoja Quebrada Salitrosa, III Región de Atacama: Santiago, Chile, Servicio Nacional de Geología y Minería, Mapas Geológicos 4, scale 1:100.000*.

- Lawrence, S., Giles, C. L., & Tsoi, A. C. (1997). Lessons in neural network training: Overfitting may be harder than expected. In AAAI/IAAI (pp. 540-545).
- Lemenkova, P. (2019). Regression models by Gretl and R statistical packages for data analysis in marine geology. *International Journal of Environmental Trends (IJENT)*, 3(1), 39-59.
- Lipin, B. R., & McKay, G. A. (Eds.). (2018). *Geochemistry and mineralogy of rare earth elements* (Vol. 21). Walter de Gruyter GmbH & Co KG.
- Liu, F., & Deng, Y. (2020). Determine the number of unknown targets in open world based on elbow method. *IEEE Transactions on Fuzzy Systems*, 29(5), 986-995.
- Long, K. R. (2011). The future of rare earth elements-will these high-tech industry elements continue in short supply. *US Geological Survey Open-File Report*, 1189, 41.
- Long, K. R., Van Gosen, B. S., Foley, N. K., & Cordier, D. (2012). The principal rare earth elements deposits of the United States: a summary of domestic deposits and a global perspective. In *Non-renewable resource issues* (pp. 131-155). Springer, Dordrecht.
- López Araya, A. B. (2017). Caracterización geoquímica, litológica y espacial de las tierras raras en el yacimiento El Teniente. [Título profesional, Universidad de Chile]. Repositorio Académico de la Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/146539>
- Lopez, G. P., Hitzman, M. W. & Nelson, E. P. (2014). Alteration patterns and structural controls of the El Espino IOCG mining district, Chile. *Miner. Depos.* 49, 235–259.
- Loyola, N. I. (2016). Origen de la mineralización y alteración hidrotermal del depósito tipo IOCG Diego de Almagro, III región de Atacama, Chile.
- Lu, J., Chen, W., Ying, Y., Jiang, S., & Zhao, K. (2021). Apatite texture and trace element chemistry of carbonatite-related REE deposits in China: Implications for petrogenesis. *Lithos*, 398, 106276.
- MacKay, D. J. (1996). Hyperparameters: optimize, or integrate out? In *Maximum entropy and bayesian methods* (pp. 43-59). Springer, Dordrecht.
- MacKay, D. J. (2003). *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, p. 640.

- MacQueen, J. (1967, June). Classification and analysis of multivariate observations. In 5th Berkeley Symp. Math. Statist. Probability (pp. 281-297). Los Angeles LA USA: University of California.
- Maksaev, V., Zentilli, M. (2002). Chilean strata-bound Cu–(Ag) deposits: an overview. In: Porter, T.M. (Ed.), Hydrothermal Iron Oxide Copper–Gold and Related Deposits: A Global Perspective. PGC Publishing, Adelaide, pp. 163–184.
- Mancheri, N. A. (2015). World trade in rare earths, Chinese export restrictions, and implications. *Resources Policy*, 46, 262-271.
- Mariano, A. N., & Mariano Jr, A. (2012). Rare earth mining and exploration in North America. *Elements*, 8(5), 369-376.
- Mariano, A.N., (1989). Economic geology of rare earth minerals, chap. 11 of Lipin, B.R., and McKay, G.A., eds., *Geochemistry and mineralogy of rare earth elements*: Washington, D.C., The Mineralogical Society of America Reviews in Mineralogy series, v. 21, p. 309–337.
- Marill, T., & Green, D. M. (1960). Statistical recognition functions and the design of pattern recognizers. *IRE Transactions on Electronic Computers*, (4), 472-477.
- Mark G, Oliver N., Williams P.J, Valenta R., & Crookes R. (2000). The evolution of the Ernest Henry Fe-oxide–(Cu–Au) hydrothermal system. In: Porter TM (ed) *Hydrothermal iron oxide copper–gold and related deposits: a global perspective*, Australian Mineral Foundation, Adelaide, pp 123–136.
- Mark, G., Oliver, N. H., & Williams, P. J. (2006). Mineralogical and chemical evolution of the Ernest Henry Fe oxide–Cu–Au ore system, Cloncurry district, northwest Queensland, Australia. *Mineralium Deposita*, 40(8), 769.
- Marschik, R. (2000). La Candelaria and the Punta del Cobre district, Chile: Early Cretaceous iron-oxide Cu-Au (-Zn-Ag) mineralization. *Hydrothermal iron oxide copper-gold and related deposits: A global perspective*, 1, 163-175.

- Marschik, R., & Fontboté, L. (1996). The Punta del Cobre belt, Northern Chile: intrusion-related mid-cretaceous Cu (-Fe) mineralization. *Géodynamique andine= Andean geodynamics= Geodinámica andina: résumés étendus= extended abstracts= resúmenes ampliados*, 687-690.
- Marschik, R., & Fontboté, L. (2001a). The Candelaria-Punta del Cobre iron oxide Cu-Au (-Zn-Ag) deposits, Chile. *Economic Geology*, 96(8), 1799-1826.
- Marschik, R., & Fontboté, L. (2001b). The Punta del Cobre Formation, Punta del Cobre–Candelaria area, northern Chile. *Journal of South American Earth Sciences*, 14(4), 401-433.
- Marschik, R., & Söllner, F. (2006). Early Cretaceous U-Pb zircon ages for the Copiapo plutonic complex and implications for the IOCG mineralization at Candelaria, Atacama region, Chile: *Mineralium Deposita*, v. 41, p. 785–801.
- Marschik, R., Fontignie, D., Chiaradia, M., & Voldet, P. (2003). Geochemical and Sr-Nd-Pb-O isotope composition of granitoids of the Early Cretaceous Copiapó plutonic complex (27°30'S), Chile. *Journal of South American Earth Sciences*, 16(5), 381–398. [https://doi.org/10.1016/S0895-9811\(03\)00104-4](https://doi.org/10.1016/S0895-9811(03)00104-4)
- Marschik, R., Fontignie, D., Chiaradia, M., & Voldet, P. (2003). Geochemical and Sr–Nd–Pb–O isotope composition of granitoids of the Early Cretaceous Copiapó plutonic complex (27° 30' S), Chile. *Journal of South American Earth Sciences*, 16(5), 381-398.
- Marschik, R., Singer, B. S., Munizaga, F., Tassinari, C., Moritz, R., & Fontboté, L. (1997). Age of Cu (-Fe)-Au mineralization and thermal evolution of the Punta del Cobre district, Chile. *Mineralium Deposita*, 32(6), 531-546.
- Marsland, S. (2009). *Machine Learning: An Algorithmic Perspective*, Machine Learning and Pattern Recognition Series, Chapman & Hall/CRC Press, Boca Raton, Florida, p. 406.
- Martin, W., Díaz, R., Núñez, R., Olivares, R., Calderón, C., & Calderón, P. (1997). The updated Candelaria geologic mine model. *Proc 8th Chilean Geologic Congr*, Antofagasta, Chile, 2, 1063-1067.
- Massari, S., & Ruberti, M. (2013). Rare earth elements as critical raw materials: Focus on international markets and future strategies. *Resources Policy*, 38(1), 36-43.

- Mathur R, Marschik R, Ruiz J, Munizaga F, Leveille RA, & Martin W. (2002). Age of mineralization of the Candelaria Fe oxide Cu- Au deposit and the origin of the Chilean iron belt, based on Re- Os isotopes. *Econ Geol* 97:59–71.
- Mauger, A. J., Ehrig, K., Kontonikas-Charos, A., Ciobanu, C. L., Cook, N. J., & Kamenetsky, V. S. (2016). Alteration at the Olympic Dam IOCG–U deposit: insights into distal to proximal feldspar and phyllosilicate chemistry from infrared reflectance spectroscopy. *Australian Journal of Earth Sciences*, 63(8), 959-972.
- Mavrogenes, J. A., & O'Neill, H. S. C. (1999). The relative effects of pressure, temperature and oxygen fugacity on the solubility of sulfide in mafic magmas. *Geochimica et Cosmochimica Acta*, 63(7-8), 1173-1180.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 27(4), 12-12.
- McGill, I. (1997). Rare earth metals, in Habashi, F., ed., *Handbook of extractive metallurgy: weinheim*, New York, Wiley-VCH, v. 3, p: 1695-1741.
- MCH. (2015, 4 noviembre). La apuesta por las tierras raras de Minera Biolantánidos. *Minería Chilena*. Recuperado 8 de febrero de 2022, de <https://www.mch.cl/entrevistas/la-apuesta-por-las-tierras-raras-de-minera-biolantanidos/#>
- MCH. (2020, 23 septiembre). Ministro Prokurica destacó potencial de Chile para producir Tierras Raras. *Minería Chilena*. Recuperado 2 de febrero de 2022, de <https://www.mch.cl/2020/09/22/ministro-prokurica-destaco-potencial-de-chile-para-producir-tierras-raras/>
- McKay, G. A. (1989). Partitioning of rare earth elements between major silicate minerals and basaltic melts. *Reviews in Mineralogy and Geochemistry*, 21(1), 45-77.
- McKeown, F. A., & Klemic, H. (1954). Rare-earth-bearing apatite at Mineville, Essex County, New York. US Department of the Interior, Geological Survey.
- McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. *Python for high performance and scientific computing*, 14(9), 1-9.

- McLennan, S. M. (2012). Geology, geochemistry, and natural abundances of the rare earth elements. In: Atwood DA (ed) The rare earth elements—fundamentals and applications. Wiley, New York, pp 1–19.
- Meilă, M. (2005). Comparing clusterings: an axiomatic view. In Proceedings of the 22nd international conference on Machine learning (pp. 577-584).
- Ménard, J. (1986). Un modèle métasomatique pour les gisements de la Ceinture de fer du Chili. *Comptes Rendus, Acad. Sci.* 302:775–779.
- Ménard, J. (1995). Relationship between altered pyroxene diorite and the magnetite mineralization in the Chilean Iron Belt, with emphasis on the El Algarrobo iron deposits (Atacama region, Chile). *Mineral. Deposita* 30, 268-274.
- Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning, neural and statistical classification. Ellis Horwood, New York, pp. 1-5.
- Miesch, A. T. (1980). Scaling variables and interpretation of eigenvalues in principal component analysis of geologic data. *Journal of the International Association for Mathematical Geology*, 12(6), 523-538.
- Migdisov, A. A., Bychkov, A. Y., Williams-Jones, A. E., & Van Hinsberg, V. J. (2014). A predictive model for the transport of copper by HCl-bearing water vapour in ore-forming magmatic-hydrothermal systems: Implications for copper porphyry ore formation. *Geochimica et Cosmochimica Acta*, 129, 33-53.
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50, 159-179.
- Mindat (2022, 15 de enero). Bayan Obo deposit (Bayun-Obo deposit; Baiyunebo deposit), Bayan Obo mining district, Baotou City (Baotou Prefecture), Inner Mongolia, China. Mindat.org. Recuperado 7 de febrero de 2022, de <https://mindat.org/loc-720.html>
- Mirkin, B. (2012). Clustering: A Data Recovery Approach, Vol. 19.
- Mirzababaei, G., Yazdi, M., Behzadi, M., & Rezvanianzadeh, M. R. (2021). REE-Th mineralization in the Se-Chahun magnetite-apatite ore deposit, central Iran: Interplay of magmatic and metasomatic processes. *Ore Geology Reviews*, 139, 104426.

- Mitchell, T. M. (1997). Machine Learning. New York: McGraw-Hill Companies Inc.
- Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*, 42(11), 30-36.
- Mitchell, T. M., & Lumsden, S. (2018). Reinforcement learning. *Machine Learning*.
- Möller, P., Cerny, P., & Saupé, F. (Eds.). (2013). Lanthanides, Tantalum and Niobium: Mineralogy, Geochemistry, Characteristics of Primary Ore Deposits, Prospecting, Processing and Applications Proceedings of a workshop in Berlin, November 1986 (Vol. 7). Springer Science & Business Media.
- Montreuil, J. F., Corriveau, L., & Grunsky, E. C. (2013). Compositional data analysis of hydrothermal alteration in IOCG systems, Great Bear magmatic zone, Canada: to each alteration type its own geochemical signature.
- Montreuil, J. F., Potter, E. G., Corriveau, L., & Davis, W. J. (2016). Element mobility patterns in magnetite-group IOCG systems: The Fab IOCG system, Northwest Territories, Canada. *Ore Geology Reviews*, 72, 562-584.
- Moradi, F., Olovsson, T., & Tsigas, P. (2012). An Evaluation of Community Detection Algorithms on Large-Scale Email Traffic. *SEA*, 7276, 283-294.
- Morales, F. N., (2022). Mineralogía y geoquímica de los óxidos de hierro en el distrito Punta del Cobre. Tesis, Universidad de Atacama. Repositorio Académico de la Universidad de Atacama. <https://repositorioacademico.uda.cl/handle/20.500.12740/16235>
- Morales, H. A. (2017). Estudio de la paragénesis y química mineral en magnetitas del depósito tipo IOCG Dominga, IV Región, Chile. Tesis, Universidad de Chile. Repositorio Académico de la Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/146729>.
- Moss, R., Scott, S. D., & Binns, R. A. (2001). Gold content of eastern Manus basin volcanic rocks: Implications for enrichment in associated hydrothermal precipitates. *Economic Geology*, 96(1), 91-107.
- Mpodozis, C., & Ramos, V. (1990). The Andes of Chile and Argentina, in Ericksen, G.E., Cañas-Pinochet, M.T., and Reinemund, J.A., eds., *Geology of the Andes and its relation to*

- hydrocarbon and mineral resources: Houston, Texas, Circum-Pacific Council for Energy and Mineral Resources, Earth Sciences Series, v. 11, p. 59–91.
- Mudd, G. M., & Jowitt, S. M. (2018). Growing global copper resources, reserves and production: Discovery is not the only control on supply. *Economic Geology*, 113(6), 1235-1267.
- Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, 24(3), 69-71.
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc."
- Munita, I (2021, 22 marzo). «Tierras raras»: Qué son, para qué se utilizan y el potencial de Chile para abrirse un espacio en este mercado. *Emol*. Recuperado 31 de enero de 2022, de <https://www.emol.com/noticias/Economia/2021/03/22/1015435/Tierras-raras-Chile.html>.
- Munizaga, R., & Lagos, M. (2015). Antecedentes geológicos del yacimiento de magnetita-apatito Los Colorados, provincia del Huasco, Tercera Region de Atacama, Chile. In *Abstracts XIV Congreso Geológico Chileno, La Serena* (pp. 102-105).
- Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), 86-97.
- Murty, M. N., & Devi, V. S. (2011). *Pattern recognition: An algorithmic approach*. Springer Science & Business Media.
- Musa, R. M., Taha, Z., Majeed, A. P. A., & Abdullah, M. R. (2019). *Machine learning in sports: identifying potential archers*. Springer Singapore.
- Na, S., Xumin, L., & Yong, G. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *2010 Third International Symposium on intelligent information technology and security informatics* (pp. 63-67). Ieee.
- Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method. In *Journal of Physics: Conference Series* (Vol. 1361, No. 1, p. 012015). IOP Publishing.

- Naranjo, J.A. (1987). Interpretación de la actividad cenozoica superior a lo largo de la Zona de Falla Atacama, Norte de Chile. *Andean Geol.* 31, 43–55. ore deposits in the Andes. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 353–364.
- Naslund, H. R., Henriquez, F., Nystroem, J. O., Vivallo, W., Dobbs, F. M., (2002). *Magmatic Iron Ores and Associated Mineralization: Examples from the Chilean High Andes and Coastal Cordillera: Hydrothermal Iron Oxide Copper-Gold and Related Deposits: A Global Perspective*, Adelaide, Australia. PGC Publishing, pp. 207–226.
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*. b, 4, 51-62. *Neural Networks*, vol. 16, no. 3, pp. 645-678.
- Nathwani, C. L., Wilkinson, J. J., Fry, G., Armstrong, R. N., Smith, D. J., & Ihlenfeld, C. (2022). Machine learning for geochemical exploration: classifying metallogenic fertility in arc magmas and insights into porphyry copper deposit formation. *Mineralium Deposita*, 57(7), 1143-1166.
- Neumann, E. R., Svensen, H. H., Polozov, A. G., & Hammer, Ø. (2017). Formation of Si-Al-Mg-Ca-rich zoned magnetite in an end-Permian phreatomagmatic pipe in the Tunguska Basin, East Siberia. *Mineralium Deposita*, 52(8), 1205-1222.
- Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann.
- Nilsson, N.J. (1965). *Learning machines*. New York: McGraw-Hill. Published in: *Journal of IEEE Transactions on Information Theory* Volume 12 Issue 3, 1966. doi: 10.1109/TIT.1966.1053912 pp. 407 – 407.
- Norton, D. L., (1982). Fluid and heat transport phenomena typical of copper-bearing pluton environments. In: S.R. Titley (Editor), *Advances in Geology of Porphyry Copper Deposits, Southwestern North America*. Univ. Arizona Press, Tuscon, pp. 59-72.
- Nystroem, J. O., & Henriquez, F., (1994). Magmatic features of iron ores of the Kiruna type in Chile and Sweden; ore textures and magnetite geochemistry. *Econ. Geol.* 89 (4), 820–839. <https://doi.org/10.2113/gsecongeo.89.4.820>.
- Oberreuter, B. (2017). Distribución de la alteración hidrotermal y su relación con mineralización de cobre en el Área de Chivato, Región de Atacama, Norte de Chile.

- Oliveros, V., Vásquez, P., Creixell, C., Lucassen, F., Ducea, M. N., Ciocca, I., Gonzáles, J., Espinoza, M., Salazar, E., Coloma, F., & Kasemann, S. A. (2020). Lithospheric evolution of the Pre-and Early Andean convergent margin, Chile. *Gondwana Research*, 80, 202-227.
- Oreskes, N. & Einaudi, M.T., (1992). Origin of hydrothermal fluids at Olympic Dam: Preliminary results from fluid inclusions and stable isotopes. *Econ. Geol.*, 64-90.
- Oreskes, N. (1990). American geological practice: participation and examination. Part I. Origin of REE-enriched hematite breccias at Olympic Dam, South Australia. Ph.D. dissertation, Stanford University, Stanford, CA, unpubl.
- Oreskes, N., & Einaudi, M. T. (1990). Origin of rare earth element-enriched hematite breccias at the Olympic Dam Cu-U-Au-Ag deposit, Roxby Downs, South Australia. *Economic Geology*, 85(1), 1-28.
- Ortiz, F., Araya, R., Franqueza, F., Moraga, A., & Zentilli, M. (1966). Informe geológico del distrito minero de Punta del Cobre: Instituto de Investigaciones Geológicas, Santiago de Chile. Special Publication, (1), 4.
- Oyarzún, J. (2007). El Modelo IOCG y el Potencial de Exploración Cuprífera de la Cordillera de la Costa del Norte de Chile.
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.
- Palma, G., Reich, M., Barra, F., Ovalle, J. T., Del Real, I., & Simon, A. C. (2021). Thermal evolution of Andean iron oxide–apatite (IOA) deposits as revealed by magnetite thermometry. *Scientific Reports*, 11(1), 1-9.
- Panchal, G., Ganatra, A., Shah, P., & Panchal, D. (2011). Determination of over-learning and over-fitting problem in back propagation neural network. *International Journal on Soft Computing*, 2(2), 40-51.
- Papangelakis, V. G., & Moldoveanu, G. (2014). Recovery of rare earth elements from clay minerals. In *Proceedings of the 1st Rare Earth Resources Conference, Milos* (pp. 191-202).

- Parak, T. (1973). Rare earths in the apatite iron ores of Lappland together with some data about the Sr, Th, and U content of these ores. Luossavaara-Kiirunavaara AB, Kiruna, Sweden.
- Park, C. F., (1961). A magnetite “flow” in northern Chile. *Econ. Geol.* 56, 431–436. <https://doi.org/10.2113/gsecongeo.56.2.431>.
- Peng, X., Yang, H., & Zhang, J. (2018). Geology, geochemistry, and genesis of the Dashui Carlin-type gold deposit in the West Qinling orogenic belt, Gansu Province, China. *Geological Journal*, 53(3), 835-856.
- Pérez, E., Cooper, M. R., & Covacevich, V. (1990). Aptian ammonite-based age for the Pabellón Formation, Atacama Region, northern Chile. *Andean Geology*, 17(2), 181-185.
- Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), 103-119.
- Pichowiak, S. (1994). Early Jurassic to early Cretaceous magmatism in the coastal cordillera and the central depression of north Chile. In *Tectonics of the southern Central Andes* (pp. 203-217). Springer, Berlin, Heidelberg.
- Piotrowski, A. P., & Napiorkowski, J. J. (2013). A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. *Journal of Hydrology*, 476, 97-111.
- Pollard, P. J. (2000). Evidence of a magmatic fluid and metal source for Fe oxide Cu-Au mineralization. *Hydrothermal iron oxide copper-gold and related deposits: A global perspective*, 1, 27-41.
- Pollard, P. J. (2006). An intrusion-related origin for Cu–Au mineralization in iron oxide–copper–gold (IOCG) provinces. *Mineralium Deposita*, 41(2), 179-187.
- Poole, D., Mackworth, A. K., & Goebel, R. (1998). *Computational intelligence: A logical approach*. Oxford University Press.
- Pop, N., Heaman, L., Edelstein, O., Isache, C., Zentilli, M., Pecskey, Z., Valdman, Z., & Rusu, C. (2000). Geocronología de las rocas ígneas y los productos de alteración hidrotermal relacionadas con la mineralización de Cu-Fe (Au) del sector Adriana-Cárola-Cobriz

- (parte este del distrito Punta del Cobre-Candelaria), en base a dataciones U-Pb (en circón), $^{40}\text{Ar}/^{39}\text{Ar}$ y K-Ar. Actas, 2, 155-160.
- Porter, T. M. (Ed.), (2000). Hydrothermal Iron Oxide Copper-gold and Related Deposits: A Global Perspective. Australian Mineral Foundation, Adelaide, pp. 349.
- Porter, T. M., (2010). The Carrapateena iron oxide copper gold deposit, Gawler Craton, South Australia: A review. In: Porter, T.M. (Ed.), Hydrothermal Iron Oxide Coppergold and Related Deposits: A Global Perspective – Advances in the Understanding of IOCG Deposits. PGC Publishing, Adelaide, pp. 191–200.
- Prieto, A. C. (2014). Elementos críticos energéticos: potencial de desarrollo en Chile. [Título Profesional, Universidad de Chile]. Repositorio Académico de la Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/116303>
- Project Jupyter. (2016). Project Jupyter. Jupyter. Recuperado 24 de agosto de 2022, de <https://jupyter.org/about>
- Pythonista. (2020, 7 noviembre). Anaconda Python y Cona. El Pythonista. Recuperado de agosto de 2022 de <https://elpythonista.com/anaconda>
- Qui, Y., (1990). Bayan Obo mineralization related to Hercynian granite? 8th IAGOD Syrup., Program with Abstracts, Ottawa, Canada, p. A252.
- Rahman, A. T. (2017). Coal trade data clustering using K-means (case study Pt. Global Bangkit Utama). ITSMART: Jurnal Teknologi dan Informasi, 6(1), 24-31.
- Ramírez, F. (2020, 14 junio). ¿Qué es y cómo funciona la Inteligencia Artificial?
- Ray, G. E., & Lefebure, D. V. (2000). A synopsis of iron oxide±Cu±Au±P±REE deposits of the Candelaria-Kiruna-Olympic Dam family. British Columbia Ministry of Energy and Mines, Geological Fieldwork, 2000-1.
- Ray, G. E., Dick, L. A., & Porter, T. M. (2002). The Productora prospect in north-central Chile: An example of an intrusion-related Candelaria type Fe-Cu-Au hydrothermal system. Porter GeoConsultancy Publishing, Adelaide, 2, 131-151.
- Reich, M., Simon, A. C., Deditius, A., Barra, F., Chrysoulis, S., Lagas, G., Tardani, D., Knipping, J., Bilenker, L., Sánchez-Alfaro, P., Roberts, M., & Munizaga, R. (2016). Trace

- element signature of pyrite from the Los Colorados iron oxide-apatite (IOA) deposit, Chile: A missing link between Andean IOA and iron oxide copper-gold systems?. *Economic Geology*, 111(3), 743-761.
- Ribeiro, R. G., & Rios, R. (2021). Temporal gap statistic: A new internal index to validate time series clustering. *Chaos, Solitons & Fractals*, 142, 110326.
- Rich, E. & Knight, K. (1991). *Artificial Intelligence* (second edition). McGraw-Hill.
- Richards, J. P., López, G. P., Zhu, J. J., Creaser, R. A., Locock, A. J., & Mumin, A. H. (2017). Contrasting tectonic settings and sulfur contents of magmas associated with Cretaceous porphyry Cu±Mo±Au and intrusion-related iron oxide Cu-Au deposits in northern Chile. *Economic Geology*, 112(2), 295-318.
- Rieger A. A., Marschik R., & Díaz, M., (2010b). The Mantoverde district: an example of distal IOCG deposits in North Chile. In Potter TM (ed) *Hydrothermal iron oxide copper-gold and related deposits: a global perspective—advances in the understanding of IOCG deposits*. Adelaide, Australian Mineral Foundation 3:273-284.
- Rieger, A. A., Marschik, R., & Díaz, M. (2012). The evolution of the hydrothermal IOCG system in the Mantoverde district, northern Chile: New evidence from microthermometry and stable isotope geochemistry. *Mineralium Deposita*, 47(4), 359-369.
- Rieger, A. A., Marschik, R., Chiaradia, M., Spangenberg, J. E. & Díaz, M. (2009). Sulphur and Lead Isotope Characteristics of Sulphides of the Mantoverde IOCG District, Chile. *Proceedings of the Tenth Biennial SGA Meeting*, Townsville.
- Rieger, A. A., Marschik, R., Díaz, M., Hölzl, S., Chiaradia, M., Akker, B., & Spangenberg, J. E. (2010a). The hypogene iron oxide copper-gold mineralization in the Mantoverde district, northern Chile. *Economic Geology*, 105(7), 1271-1299.
- Rieuwers, M., & Caté, A. (2019). Integrating a Minerals Systems Approach with Machine Learning: A Case Study of ‘Modern Minerals Exploration’ in the Mt Woods Inlier – northern Gawler Craton, South Australia. *ASEG Extended Abstracts*, 2019(1), 1–4. <https://doi.org/10.1080/22020586.2019.12073159>

- Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, p. 403.
- Riquelme, S. (2020, 29 septiembre). En Chile existen 750 con tierras raras que no han sido extraídas. *Guía Minera de Chile*. Recuperado 31 de enero de 2022, de <https://www.guiaminera.cl/en-chile-existen-unos-750-relaves-con-tierras-raras-que-no-han-sido-extraidas/#:~:text=Las%20tierras%20raras%20tienen%2015,m%C3%A1s%20el%20itri o%20y%20 scandio.>
- Rodríguez, D. (2019 mayo). Los conceptos de sesgo y varianza en aprendizaje automáticos. *Analyticslane*. *Analyticslane*. Recuperado el 20 de noviembre de 2021 de <https://www.analyticslane.com/2019/05/24/los-conceptos-de-sesgo-y-varianza-en-aprendizaje-automaticos/>.
- Rodríguez, D. (2021, 13 junio). ¿Cuál es la diferencia entre parámetro e hiperparámetro? *Analytics Lane*. Recuperado 26 de febrero de 2022, de <https://www.analyticslane.com/2019/12/16/cual-es-la-diferencia-entre-parametro-e-hiperparametro/>
- Rodriguez-Galiano, V. F., Chica-Olmo, M., & Chica-Rivas, M. (2014). Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain. *International Journal of Geographical Information Science*, 28(7), 1336–1354. <https://doi.org/10.1080/13658816.2014.885527>
- Rodriguez-Mustafa, M. A., Simon, A. C., del Real, I., Thompson, J. F., Bilenker, L. D., Barra, F., Bindemn, I., & Cadwell, D. (2020). A continuum from iron oxide copper-gold to iron oxide-apatite deposits: Evidence from Fe and O stable isotopes and trace element chemistry of magnetite. *Economic Geology*, 115(7), 1443-1459.
- Rohwer, R., Wynne-Jones, M. & Wysotzki, F. (1994). *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, New York, pp. 84-106.
- Rojas Olmedo, J. P. (2020). Desarrollo de metodología de predicción y calibración de parámetros de simulaciones DEM aplicado a flujo de mineral de cobre basada en machine learning.

- Rojas P, Barra F, Reich M, Deditius A, Simon A, Uribe F, Romero R, Rojo M (2018). A genetic link between magnetite mineralization and diorite intrusion at the El Romeral iron oxide-apatite deposit, northern Chile. *Miner Deposita*, 53(7), 947-966. <https://doi.org/10.1007/s0>
- Rollinson, H. R., (1993). *Using Geochemical Data: Evaluation, Presentation, Interpretation*. Longman, Harlow, p. 352.
- Rolon-Mérette, D., Ross, M., Rolon-Mérette, T., & Church, K. (2016). Introduction to Anaconda and Python: Installation and setup. *Python for research in psychology*, 16(5), S5-S11.
- Rondón, I. (2022, febrero). ¿Qué es Anaconda? – Escuela Internacional de Posgrados. Másteres Online N°1 Empleabilidad. Recuperado 25 de agosto de 2022, de <https://eiposgrados.com/blog-python/que-es-anaconda/>
- Roselli, D., Matthews, J., & Talagala, N. (2019). Managing bias in AI. In *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 539-544).
- Rosenberg, A. & Hirschberg, J. (2007), "V-Measure: A conditional entropy-based external cluster evaluation measure", *Proceedings of the Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Prague, Czech Republic, pp. 410–420.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- Roskill (2018). *Rare Earths. Global Industry, Markets & Outlook 2018*.
- Rotherham, J. F., Blake, K. L., Cartwright, I., Williams, P.J. (1998). Stable isotope evidence for the origin of Meso-proterozoic Starra Au-Cu deposit, Cloncurry district, NW Queensland. *Economic Geology*, vol. 93, pp. 1435-1449.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53-65.
- Rozada, S. (2019). ¿Qué son los Hiperparámetros y para qué sirven en “machine learning”? Quora. Recuperado 26 de febrero de 2022 de <https://es.quora.com/Qu%C3%A9-son-los-hiperparametros-y-para-qu%C3%A9-sirven-en-machine-learning>

- Rudnick, R. L. (1995). Making continental crust. *Nature*, 378(6557), 571-578.
- Ruiz, C., & Ericksen, G. E. (1962). Metallogenetic provinces of Chile, S. A. *Economic Geology*, 57(1), 91–106. doi:10.2113/gsecongeo.57.1.91
- Ruiz, C., & Peebles, F. (1988). *Geología, distribución y génesis de los yacimientos metalíferos chilenos*. Editorial Universitaria. Santiago.
- Ruiz, C., Aguirre, L., Corvalán, J., Klohn, C., Klohn, E., & Levi, B. (1965). *Geología y yacimientos metalíferos de Chile*. Instituto de Investigaciones Geológicas, 305.
- Ruiz, C., Ortiz, F., Moraga, A., & Aguilar, A. (1968). Genesis of the Chilean Iron ore deposits of Mesozoic age.
- Russell, S. J., & Norvig, P. (2004). *Inteligencia Artificial: un enfoque moderno* (No. 04; Q335, R8y 2004.).
- Ryan, P. J., Lawrence, A. I., Jenkins, R. A., Matthews, J. P., Zamora, J. C., Marino, E., & Urqueta, I. (1994). The Candelaria copper–gold deposit, Chile: Congreso Geológico Chileno, 7 th, Concepción. *Actas*, 2, 1616-1617.
- Ryan, P., Lawrence, A., Jenkins, R., Matthews, J., Zamora, J.C., Marino, E., y Urqueta, I., (1995). The Candelaria copper-gold deposit, Chile. *Arizona Geological Society Digest*, 20, 625-645.
- Saaty, T., & Gass, S. (1954). Parametric objective function (part 1). *Journal of the Operations Research Society of America*, 2(3), 316-319.
- Salazar, E., Barra, F., Reich, M., Simon, A., Leisen, M., Palma, G., Romero, R., Rojo, M., & Rojo, M. (2020). Trace element geochemistry of magnetite from the Cerro Negro Norte iron oxide–apatite deposit, northern Chile. *Mineralium Deposita*, 55(3), 409-428.
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.
- Santos, P.R. (2022, 24 enero). Datos de entrenamiento vs datos de test – Think Big Empresas. Think Big. Recuperado 24 de febrero de 2022, de <https://empresas.blogthinkbig.com/datos-entrenamiento-vs-datos-de-test/>

- Sarparandeh, M., & Hezarkhani, A. (2016). Application of Self-Organizing Map for Exploration of REEs' Deposition. *Open Journal of Geology*, 06(07), 571–582. <https://doi.org/10.4236/ojg.2016.67045>
- Sarparandeh, M., & Hezarkhani, A. (2017). Studying distribution of rare earth elements by classifiers, Se-Chahun iron ore, Central Iran. *Acta Geochimica*, 36(2), 232-239.
- Schacht, U., Wallmann, K., & Kutterolf, S. (2010). The influence of volcanic ash alteration on the REE composition of marine pore waters. *Journal of Geochemical Exploration*, 106(1-3), 176-187.
- Scheuber, E., & Andriessen, P.A.M. (1990). The kinematic and geodynamic significance of the Atacama fault zone, northern Chile: *Journal of Structural Geology*, v. 12, p. 243–257.
- Scheuber, E., & Gonzalez, G. (1999). Tectonics of the Jurassic-Early Cretaceous magmatic arc of the north Chilean Coastal Cordillera (22°–26° S): a story of crustal deformation along a convergent plate boundary. *Tectonics* 18, 895–910.
- Schmandt, D. S., Cook, N. J., Ciobanu, C. L., Ehrig, K., Wade, B. P., Gilbert, S., & Kamenetsky, V. S. (2019). Rare earth element phosphate minerals from the Olympic Dam Cu-U-Au-Ag deposit, South Australia: Recognizing temporal-spatial Controls on REE mineralogy in an evolved IOCG system. *The Canadian Mineralogist*, 57(1), 3-24.
- Schölkopf, B. (2003). Statistical learning theory, capacity, and complexity. *Complexity*, 8(4), 87-94. Science & Business Media.
- Schulz, K. J., DeYoung, J. H., Seal, R. R., & Bradley, D. C. (Eds.). (2018). Critical mineral resources of the United States: economic and environmental geology and prospects for future supply. Geological Survey.
- Seeger, C. M., Nuelle, L. M., Sidder, G. B., Marikos, M. A., & Smith, D. C. (2001). Geologic maps and cross sections of mine levels at the Pea Ridge iron mine, Washington County, Missouri (No. 2353).
- Segerstrom, K. (1960). Cuadrángulo Quebrada Paipote: Provincia de Atacama. *Carta Geológica de Chile*, 2(1). Instituto de Investigaciones Geológicas, Santiago de Chile 35 p.

- Segerstrom, K., & Parker, R. L. (1959). Cuadrángulo Cerrillos, Provincia de Atacama: Santiago, Instituto de Investigaciones Geológicas, Carta Geológica de Chile, v. 1, no. 2, 33 p.
- Segerstrom, K., & Ruiz, C. (1962). Geología del Cuadrángulo Copiapó, Provincia de Atacama. Instituto de Investigaciones Geológicas, Santiago, Carta Geológica de Chile, 1, no. 3, 1:50.000.
- Sepúlveda, P., & Naranjo, J. A. (1982). Geología de la Hoja Carrera Pinto. Escala 1:100.000. Servicio Nacional Geología y Minería, Carta Geológica de Chile, 53, 1-62.
- SERNAGEOMIN (2003). Mapa Geológico de Chile escala 1:1.000.000. Servicio Nacional de Geología y Minería.
- SERNAGEOMIN. (2014). Anomalías geoquímicas en el norte grande y norte chico. Plan nacional de geología, unidad de geoquímica del gobierno de Chile.
- SERNAGEOMIN. (2020, 3 de diciembre). Plan Nacional Geología. Geoquímica. Recuperado 27 de enero de 2022, de <https://plannacionalgeologia.sernageomin.cl/geoquimica/>
- Shikazono, N., Ogawa, Y., Utada, M., Ishiyama, D., Mizuta, T., Ishikawa, N., & Kubota, Y. (2008). Geochemical behavior of rare earth elements in hydrothermally altered rocks of the Kuroko mining area, Japan. *Journal of Geochemical Exploration*, 98(3), 65-79.
- Sidyakov, M. (2022, 17 marzo). Calinski-Harabasz Index for K-Means Clustering Evaluation using Python. PyShark. Recuperado el 22 de febrero de 2023 de <https://pyshark.com/calinski-harabasz-index-for-k-means-clustering-evaluation-using-python/>
- Siless, V., Medina, S., Varoquaux, G., & Thirion, B. (2013). A comparison of metrics and algorithms for fiber clustering. In 2013 International Workshop on Pattern Recognition in Neuroimaging (pp. 190-193). IEEE.
- Sillitoe, R. H. (2003). Iron oxide-copper-gold deposits: an Andean view. *Mineralium Deposita*, 38(7), 787-812.
- Simandl, G. J. (2014). Geology and market-dependent significance of rare earth element resources. *Mineralium Deposita*, 49(8), 889-904.

- Simon, A. C., Pettke, T., Candela, P. A., Piccoli, P. M., & Heinrich, C. A. (2004). Magnetite solubility and iron transport in magmatic-hydrothermal environments. *Geochimica et Cosmochimica Acta*, 68(23), 4905-4914.
- Sivia, D.S. (1996). *Data Analysis: A Bayesian Tutorial*, Oxford University Press, New York, p. 189.
- Skirrow, R. (2014). The Olympic IOCG province (Gawler craton): Lithospheric-to district-scale controls on ore formation, and targeting of IOCG mineral systems. In *Proceedings of the Geoconferences IOCG Workshop, Perth, Australia (Vol. 29)*.
- Skirrow, R. G., Bastrakov, E. N., Barovich, K., Fraser, G.L., Creaser, R.A., Fanning, C.M., Raymond, O.L., Davidson, G.J., (2007). Timing of iron oxide Cu-Au-(U) hydrothermal activity and Nd isotope constraints on metal sources in the Gawler craton, South Australia. *Econ. Geol.* 102 (8), 1441–1470.
- Skirrow, R. G., Murr, J., Schofield, A., Huston, D. L., van der Wielen, S., Czarnota, K., Coghlan, R., Highet, L., Connolly, D., Doublier, M., & Duan, J. (2019). Mapping iron oxide Cu-Au (IOCG) mineral potential in Australia using a knowledge-driven mineral systems-based approach. *Ore Geology Reviews*, 113, 103011.
- Skolnick, J., Gao, M., Zhou, H., & Singh, S. (2021). AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. *Journal of Chemical Information and Modeling*.
- Sleep, N. H., (1983). Hydrothermal convection at ridge axes. In: P.A. Rona, K. Bostrom, L. Laubier and K.L. Smith Jr. (Editors), *Hydrothermal Processes at Seafloor Spreading Centers*. NATO Conf. Ser., IV, Marine Sciences, 12: 71-82.
- Sneath, P. H., & Sokal, R. R. (1973). *Numerical taxonomy. The principles and practice of numerical classification*.
- Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, 33-40.
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In *AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence*, Hobart,

- Australia, December 4-8, 2006. Proceedings 19 (pp. 1015-1021). Springer Berlin Heidelberg.
- Solís, I. N., (2022). Cristaloquímica de sulfuros de fierro y fierro-cobre en el distrito Punta del Cobre: Implicancias para el control mineralógico de cobalto y otros elementos estratégicos. [Título Profesional, Universidad de Atacama].
- Stahl, D., & Sallis, H. (2012). Model-based cluster analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(4), 341-358.
- Stegen, K. S. (2015). Heavy rare earths, permanent magnets, and renewable energies: An imminent crisis. *Energy Policy*, 79, 1-8.
- Stern, C.R., Mohseni, P.P., & Fuenzalida, P.R. (1991). Petrochemistry and tectonic significance of Lower Cretaceous Barros Arana Formation basalts, southernmost Chilean Andes. *J. S. Am. Earth Sci.* 4, 331–342.
- Stosch, H. G., Romer, R. L., Daliran, F., & Rhede, D. (2011). Uranium–lead ages of apatite from iron oxide ores of the Bafq District, East-Central Iran. *Mineralium Deposita*, 46(1), 9-21.
- Suárez, M., & Bell, C. M. (1987). Upper Triassic to Lower Cretaceous continental and coastal saline lake evaporites in the Atacama region of northern Chile. *Geological Magazine*, 124(5), 467-475.
- Sun, J. G., Liu, J., & Zhao, L. Y. (2008). Clustering algorithms research. *Journal of software*, 19(1), 48-61.
- Sun, S.S., & McDonough, W.F., (1989). Chemical and isotopic characteristics of ocean basalts: Implications for mantle composition and processes: *Geological Society of London Special Publication* 42, p. 313–345.
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP conference series: materials science and engineering* (Vol. 336, No. 1, p. 012017). IOP Publishing.

- Tappert, M. C., Rivard, B., Giles, D., Tappert, R., & Mauger, A. (2013). The mineral chemistry, near-infrared, and mid-infrared reflectance spectroscopy of phengite from the Olympic Dam IOCG deposit, South Australia. *Ore Geology Reviews*, 53, 26-38.
- Tardani, D., Reich, M., Deditius, A. P., Chryssoulis, S., Sánchez-Alfaro, P., Wrage, J., & Roberts, M. P. (2017). Copper–arsenic decoupling in an active geothermal system: A link between pyrite and fluid composition. *Geochimica et Cosmochimica Acta*, 204, 179-204.
- Taylor, S. R., & McLennan, S. M. (1985). The continental crust: its composition and evolution.
- Thanaki, J. (2017). Python natural language processing. Packt Publishing Ltd.
- Thiele, R., & Pincheira, M. (1987). Tectonica transpresiva v movimiento de desgarre en el segmento sur de la Zona de Falla Atacama, Chile. *Andean Geology*, (31), 77-94.
- Thomson, J., Higgs, N. C., Croudace, I. W., Colley, S., & Hydes, D. J. (1993). Redox zonation of elements at an oxic/post-oxic boundary in deep-sea sediments. *Geochimica et Cosmochimica Acta*, 57(3), 579-595.
- Thorndike, R. (1953). Who belongs in the family?. *Psychometrika*, 18(4), 267-276.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423.
- Torab, F. M., & Lehmann, B. (2007). Magnetite-apatite deposits of the Bafq district, Central Iran: apatite geochemistry and monazite geochronology. *Mineralogical magazine*, 71(3), 347-363.
- Torralba, A., & Efros, A. A. (2011). Unbiased look at dataset bias. In *CVPR 2011* (pp. 1521-1528). IEEE.
- Tosi, S. (2009). Matplotlib for Python developers. Packt Publishing Ltd.
- Treloar, P. J., & Colley, H. (1996). Variations in F and Cl contents in apatites from magnetite—apatite ores in northern Chile, and their ore-genetic implications. *Mineralogical Magazine*, 60(399), 285-301.

- Tsay, A., Zajacz, Z., & Sanchez-Valle, C. (2014). Efficient mobilization and fractionation of rare-earth elements by aqueous fluids upon slab dehydration. *Earth and Planetary Science Letters*, 398, 101-112.
- U.S. Geological Survey, (2005). Mineral Commodity Summaries.
- U.S. Geological Survey, (2015). Mineral Commodity Summaries.
- Ullrich, T. D., Clark, A. H., & Stanley, C. J. (1999). The Candelaria copper-gold deposit, Region III, Chile: Paragenesis, geochronology and fluid composition. *Mineral deposits: Processes to processing*: Rotterdam, Balkema, 201-204.
- Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, 109(2), 373–440. <https://doi.org/10.1007/s10994-019-05855-6>.
- Van Gosen, B. S., Verplanck, P. L., Seal II, R. R., Long, K. R., & Gambogi, J. (2017). Rare-earth elements (No. 1802-O). US Geological Survey.
- Van Rijsbergen, C. (1979, September). Information retrieval: theory and practice. In *Proceedings of the joint IBM/University of Newcastle upon tyne seminar on data base systems* (Vol. 79).
- Vapnik V.N. & A. Ya. Lerner. (1963). Recognition of Patterns with help of Generalized Portraits, *Avtomat. iTele-mekh.*, 774-780.
- Vapnik, V. & Chervonenkis, A. (1964). On a class of perceptrons. *Automat. Remote Control* 25, 103-109.
- Varela, J. P. (2021). Mineralización y alteración hidrotermal del yacimiento La Estrella, Distrito El Zapallo, Región de Atacama. [Título Profesional, Universidad de Chile]. Repositorio Académico de la Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/182327>
- Vega, F. (2021, 29 octubre). ¿Chile, futuro proveedor de tierras raras? Los preciados metales del mundo tecnológico están en penca. *DFmas*. Recuperado 31 de enero de 2022, de <https://dfmas.df.cl/df-mas/por-dentro/chile-futuro-proveedor-de-tierras-raras-los-preciosos-metales-del>
- Velasco, F., & Tornos, F. (2009). Origen de las pegmatitas de magnetita--apatito de la Mina Carmen del Fierro (Atacama, Chile). *Resumen SEM, Macla* n°11.

- Velasco, F., Tornos, F., Hanchar, J.M., (2016). Immiscible iron-and silica-rich melts and magnetite geochemistry at the El Laco volcano (northern Chile): evidence for a magmatic origin for the magnetite deposits. *Ore Geol. Rev.* 79, 346–366. <https://doi.org/10.1016/j.oregeorev.2016.06.007>.
- Velásquez, G., Carrizo, D., Salvi, S., Vela, I., Pablo, M., & Pérez, A. (2020). Tracking cobalt, REE and gold from a porphyry-type deposit by LA-ICP-MS: A geological approach towards metal-selective mining in tailings. *Minerals*, 10(2), 109.
- Veloso, E. A., Heuser, G., Arancibia, G., & Cordeiro, P. (2019). Trace elements distribution in magnetite: the relationship between tectonics and hydrothermal fluid flow in the Dominga Fe-Cu deposit, Coastal Cordillera, Northern Chile. In *AGU Fall Meeting Abstracts* (Vol. 2019, pp. V33C-0259).
- Veloso, E., Cembrano, J., Arancibia, G., Heuser, G., Neira, S., Siña, A., Garrido, I., Vermeesch, P. & Selby, D. (2017). Tectono-metallogenetic evolution of the Fe–Cu deposit of Dominga, northern Chile. *Mineralium Deposita*, 52(4), 595-620.
- Venegas, D. A. (2015). Caracterización, génesis y evolución de Mina Maciso, depósito pegmatítico tipo tierras raras, Cordillera de la Costa, Región de Atacama, Chile. [Título Profesional, Universidad de Atacama]. Repositorio Académico de la Universidad de Atacama. <https://repositorioacademico.uda.cl/handle/20.500.12740/8299>
- Venners, B. (2003, 13 enero). Artima – The Making of Python. Artima. Recuperado 23 de Agosto de 2022, de <https://www.artima.com/articles/the-making-of-python>
- Verdugo-Ihl, M. R., Ciobanu, C. L., Courtney-Davies, L., Cook, N. J., Slattery, A., Ehrig, K., Tornos, F., & Hanchar, J. M. (2022). U-Pb Geochronology and mineralogy of hematite from Mantoverde and Carmen de Cobre, Northern Chile: Constraints on Andean IOCG Mineralization. *Economic Geology*, 117(4), 943-960.
- Verplanck, P. L., Nordstrom, D. K., Taylor, H. E., & Kimball, B. A. (2004). Rare earth element partitioning between hydrous ferric oxides and acid mine water during iron oxidation. *Applied Geochemistry*, 19(8), 1339-1354.

- Verplanck, P.L., Van Gosen, B.S., Seal, R.R., & McCafferty, A.E., (2014). A deposit model for carbonatite and peralkaline intrusion-related rare earth element deposits: U.S. Geological Survey Scientific Investigations Report 2010 –5070 –J, 58 p.
- Vidal C, C. E., Injoke-Espinoza, J., Sidder, G. B., & Mukasa, S. B. (1990). Amphibolitic Cu-Fe skarn deposits in the central coast of Peru. *Economic Geology*, 85(7), 1447-1461.
- Vila, T., Richard, Z., & Lindsay, N. (1996). Geology of the Manto Verde copper deposit, Northern Chile: a specularite-rich, hydrothermal-tectonic breccia related to the Atacama Fault Zone.
- Vinh, N. X., Epps, J., & Bailey, J. (2009, June). Information theoretic measures for clusterings comparison: is a correction for chance necessary?. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1073-1080).
- Vivallo, W., Henríquez, F., & Espinoza, S. (1995). Los depósitos de hierro del tipo magnetita-apatita: geoquímica de las rocas volcánicas asociadas y potencialidad de la mena de hierro como fuente de mineralización de oro. *Andean Geology*, 22(2), 159-175.
- Voncken, J. H. L. (2016). *The rare earth elements: an introduction*. Springer International Publishing.
- Waljee, A. K., & Higgins, P. D. (2010). Machine learning in medicine: a primer for physicians. *Official journal of the American College of Gastroenterology| ACG*, 105(6), 1224-1226.
- Wang, B., & Gong, N. Z. (2018). Stealing hyperparameters in machine learning. In *2018 IEEE Symposium on Security and Privacy (SP)* (pp. 36-52). IEEE.
- Weber, R. J., & Reisman, D. J. (2012). Rare earth elements: A review of production, processing, recycling, and associated environmental issues. *US EPA Region*, 8, 189-200.
- Widrow, B. (1959). Adaptive sampled-data systems. A statistical theory of adaptation, *WESCON Conv. Rec.* 3 (4).
- Widrow, B. (1961). Statistical analysis of amplitude-quantized sampled-data systems. *Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry*, 79(6), 555-568.

- Williams, P., (2010). Classifying IOCG deposits. In: Corriveau, L. & Mumin, A.H. (eds.), Exploring for iron oxide copper-gold deposits: Canada and global analogues. Geological Association of Canada, Short Course Notes, No. 20, pp. 13–21.
- Williams, P.J., Barton, M. D., Johnson, D.A., Fontbote, L., de Haller, A., Mark, G., Oliver, N.H.S., Marschik, R., (2005). Iron oxide copper-gold deposits: geology, space-time distribution, and possible modes of origin. *Econ. Geol.* 100, 371–405.
- Williams, P.J., Skirrow, R. G., (2000). Overview of iron oxide-copper-gold deposits in the Curnamona Province and Cloncurry district (eastern Mount Isa Block), Australia. In: Porter, T.M. (Ed.), *Hydrothermal Iron Oxide Copper-gold and Related Deposits: A Global Perspective*, Volume 1. Porter Geoscience Consultancy Publishing, Adelaide, pp. 105–122.
- Williams, R. (1987). A class of gradient-estimation algorithms for reinforcement learning in neural networks. In *Proceedings of the International Conference on Neural Networks* (pp. II-601).
- Williams-Jones, A. E., & Heinrich, C. A. (2005). Vapor transport of metals and the formation of magmatic-hydrothermal ore deposits: 100th Anniversary Special Paper. *Economic Geology*, 100(7), 1287-1312.
- Williams-Jones, A. E., & Migdisov, A. A. (2014). Experimental constraints on the transport and deposition of metals in ore-forming hydrothermal systems.
- Williams-Jones, A. E., Migdisov, A. A., & Samson, I. M. (2012). Hydrothermal Mobilisation of the Rare Earth Elements—a Tale of “Ceria” and “Yttria”, *Elements*, 8, 355–360.
- Wilson, J., Dallmeyer, R.D., & Grocott, J. (2000). New $^{40}\text{Ar}/^{39}\text{Ar}$ dates from the Las Tazas Complex, northern Chile: Tectonic significance: *Journal of South American Earth Sciences*, v. 13, p. 115–122.
- Winston, P. H. (1992). *Artificial Intelligence* (Third edition). Addison-Wesley.
- Witten, I.H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edn, Morgan Kaufmann Series in Data Management Systems, Elsevier/Morgan Kaufman, San Francisco, California, p. 525.

- Wood, S. A., & Samson, I. M. (1998). Solubility of ore minerals and complexation of ore metals in hydrothermal solutions.
- Wu, C., Yuan, Z., & Bai, G. (1995). Rare earth deposits in China. Mineralogical Society Series, 7, 281-310.
- Xiao, Y., & Yu, J. (2007). Gap statistic and K-means algorithm. J. Comput. Res. Dev, 44, 176-180.
- Xu, C., Wang, L., Song, W., & Wu, M. (2010). Carbonatites in China: A review for genesis and mineralization. Geoscience Frontiers, 1(1), 105-114.
- Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on neural networks, 16(3), 645-678.
- Yakubovich, V. A. (1963). Machines that learn to recognize patterns. In Metodi Vichisleniy, Vol.2. Leningrad State University Press, Leningrad. (In Russian. To be translated from the reprint: Vestnik of Saint-Petersburg University. Mathematics, 4, 2019). York, p. 188.
- Yakubovich, V. A. (1965). Certain general theoretical principles in the design of learning pattern recognition systems, Part I. In Vichislitelnaia Tekhnikai Voprosi Programirovania, Vol. 4. Leningrad State University Press, Leningrad. (In Russian).
- Yakubovich, V. A. (1966). Recurrent finitely convergent algorithms for solving systems of inequalities. Sov. Math. Doklady 1966, V. 7, pp. 300–304.
- Yalalov, D. (2022). Machine Learning. Metaverse Post. Recuperado 26 de abril de 2023, de <https://mpost.io/glossary/machine-learning/>.
- Yang, K.F., Fan, H.R., Hu, F.F., & Wang, K.Y., (2010). Intrusion sequence of carbonatite dykes and REE enrichment mechanism in Bayan Obo district, Inner Mongolia: Acta Petrologica Sinica, v. 26, no. 5, p. 1523–1529.
- Yang, K.F., Fan, H.R., Santosh, M., Hu, F.F., & Wang, K.Y., (2011). Mesoproterozoic carbonatitic magmatism in the Bayan Obo deposit, Inner Mongolia, North China: constraints for the mechanism of super accumulation of rare earth elements. Ore Geology Reviews 40, 122e131.

- Yang, X. J., Lin, A., Li, X. L., Wu, Y., Zhou, W., & Chen, Z. (2013). China's ion-adsorption rare earth resources, mining consequences and preservation. *Environmental Development*, 8, 131-136.
- Young, A., Flament, N., Maloney, K., Williams, S., Matthews, K., Zahirovic, S., & Müller, R. D. (2019). Global kinematics of tectonic plates and subduction zones since the late Paleozoic Era. *Geoscience Frontiers*, 10(3), 989-1013.
- Yu, L., Porwal, A., Holden, E. J., & Dentith, M. C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. *Computers & Geosciences*, 45, 229-239.
- Yuan, C., & Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. *J*, 2(2), 226-235.
- Zamora, R., & Castillo, B. (2001). Mineralización de Fe-Cu-Au en el Distrito Mantoverde, Cordillera de la Costa, III Región de Atacama, Chile. In *Proc. 2nd Congreso Internacional de Prospectores y Exploradores, proEXPLO, Conferencias, Instituto de Ingenieros de Minas del Perú, Lima*.
- Zandiyyeh, F., Shayestefar, M. R., Ranjbar, H., & Saadat, S. (2016). Prospectivity mapping of iron oxide-copper-gold (IOCG) deposits using support vector machine method in Feyzaabad area (east of Iran). *Journal of Himalayan Earth Sciences*, 49(2), 50-62.
- Zapata, V. A. (2020). Análisis geoquímico y geoestadístico de la concentración elemental en los suelos de las comunas de Quintero y Puchuncaví. (pp. 33-34).
- Zaremotlagh, S., & Hezarkhani, A. (2016). A geochemical modeling to predict the different concentrations of REE and their hidden patterns using several supervised learning methods: Choghart iron deposit, bafq, Iran. *Journal of Geochemical Exploration*, 165, 35–48. <https://doi.org/10.1016/j.gexplo.2016.02.001>
- Zaremotlagh, S., Hezarkhani, A., & Sadeghi, M. (2016). Detecting homogenous clusters using whole-rock chemical compositions and REE patterns: A graph-based geochemical approach. *Journal of Geochemical Exploration*, 170, 94–106. <https://doi.org/10.1016/j.gexplo.2016.08.017>

- Zeng, J. W., (1990). The sedimentary environment and metallogenetic model of Proterozoic Fe-Nb-REE deposits. In: Bai Yun O Bao, China. 8th IAGOD Symp. Program with Abstracts, Ottawa, Canada, p. A251.
- Zhang, C. H., & Huang, J. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression.
- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3), 107-115.
- Zhongxin, Y., Ge, B., Chenyu, W., Zhongqin, Z., & Xianjiang, Y. (1992). Geological features and genesis of the Bayan Obo REE ore deposit, Inner Mongolia, China. *Applied Geochemistry*, 7(5), 429-442.
- Zhu, X. (2008). Semi-supervised learning literature survey. Technical Report. 1530, University of Wisconsin- Madison.
- Zhu, X., Vondrick, C., Ramanan, D., & Fowlkes, C. C. (2012). Do We Need More Training Data or Better Models for Object Detection? In *BMVC* (Vol. 3, No. 5).

ANEXOS

ANEXO 1.- Conjunto de datos de estudio.

Muestra	Ubicación	Referencia	Método	Roca	Alteración	Este	Norte	SiO2	TiO2	Al2O3	Fe2O3	MnO	MgO	CaO	Na2O	K2O	P2O5	La	Ce	Pr	Nd	Sm	Eu	Gd	Dy	Ho	Er	Tm	Yb	Lu	Tb	Sc	Y	Sr
MV96-1 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	1	1	369000	7062500	31.71	0.6	10.44	39.53	0.04	5.61	0.23	0.16	3.92	0.23	434	692	1.4	171	21.3	3.3	1.6	1.8	0.4	1.4	0.3	0.8	0.1	0.5	20.1	8	40
MV07450 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	1	1	369717	7062195	49.73	0.99	18.32	6.67	0.11	4.28	2.58	1.37	10.45	0.17	9.4	12.4	2	5.2	1.2	0.3	3.7	4.1	0.9	2.7	0.4	1.7	0.3	0.3	19	11	198
MV96-5 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	2	1	3690000	7062500	28.66	0.55	8.73	51.44	0.03	2.31	0.04	0.12	4.42	0.04	116	217	0.7	68	10.5	1.8	0.5	0.5	0.1	0.5	0.1	1.3	0.2	0.5	24.9	4	40
MV96-5-1 (F)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	3	2	3690000	7062500	54.8	0.57	17.1	8.16	0.05	3	0.07	0.26	11.35	0.07	260	528	0	186	25	4.4	0	0	0	0	0	2.6	0.5	0.5	20.6	8	110
MV96-5-2 (Matrix)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	3	2	3690000	7062500	32.6	0.57	10.23	45.17	0.04	2.55	0.04	0.12	5.48	0.05	145	271	0	86	12.7	2.2	0	0	0	0	1.1	0.3	0.5	24.3	4	51	
MV07485 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	4	3	369544	7062396	56.9	0.18	12.43	8.09	0.23	1.76	2.06	1.34	8.67	0.06	5.8	13.1	0	8.9	2.5	0.8	0	0	0	0	2.5	0.4	0.6	5	25	34	
MV07427 (F)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	4	3	369717	7062195	69.51	0.28	13.19	1.66	0.07	0.44	1.3	0.2	10.27	0.03	3.6	6.3	0	2.7	0.6	0.2	0	0	0	0	0	0.7	0.1	0.1	4	5	90
MV07514 (F)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	4	3	369769	7062449	54.3	0.58	17.08	5.35	0.11	1.14	2.01	0.18	13.42	0.22	4.3	9.3	1.1	4.4	1	0.3	1.1	0.2	1.1	0.2	0.9	0.2	1.3	0.3	4	7	64
MV07526 (F)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	4	3	369133	7062545	56.94	0.65	16.35	6.45	0.02	1.93	0.06	0.26	12.2	0.03	2.9	5.6	0.8	2.9	0.6	0.2	0.7	0.1	0.9	0.2	0.7	0.1	1	0.2	9	4	61
MV96-4 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	5	4	3690000	7062500	51.06	0.76	16.16	11.18	0.1	5.05	0.05	0.24	10.13	0.03	42.9	75	103	23	5.3	1.5	26.4	0.5	6.3	0.8	1.8	0.2	2.1	0.4	30.3	9	108
MV07524 (F)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	5	4	369133	7062545	50.56	0.81	19.7	10.37	0.11	5.38	0.06	0.09	8.11	0.12	574	937	14.1	279	45.3	9.2	5.2	2.1	2.9	0.6	1.6	0.2	1.3	0.2	45	17	33
MV07464 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	6	5	368781	7063109	45.73	0.83	13.29	18.76	0.11	8.34	1.71	0.14	3.64	0.14	57.1	114	2.7	45.3	8.5	2	2.1	0.6	1.9	0.4	1.3	0.2	1.4	0.2	34	14	52
MV07540 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	6	6	369218	7062566	56.73	0.68	13.9	14.93	0.04	2.4	0.37	3.9	2.6	0.15	9.2	21.3	14.2	9.3	2.1	0.5	4.1	0.3	3.1	0.6	1.9	0.3	1.3	0.2	18	9	36
MV07335 (WR)	Mantoverde Sur	Rieger et al.. 2010	ICP-MS	7	7	369653	7061600	56.26	0.58	12.44	9.22	0.08	5.21	3.71	0.15	3.21	0.12	83.8	139	0	40.3	6	1.1	0	0.6	0	0	0	0	1.6	0.2	13	18	27
MV07337 (WR)	Mantoverde Sur	Rieger et al.. 2010	ICP-MS	7	8	369653	7061600	80.25	0.28	6.99	1.49	0.02	0.61	0.74	0.2	5.28	0.04	3.6	8.1	0.9	3	0.6	0.5	0.8	0.1	0.8	0.2	0.8	0.1	1	0.2	3	4	72
MV07356 (F)	Mantoverde Sur	Rieger et al.. 2010	ICP-MS	4	9	369776	7061620	38.13	0.41	13.88	24.81	0.15	3.48	0.57	0.26	6.23	0.01	2.5	5.2	0.6	2.5	0.6	0.2	0.8	0.2	1.6	0.4	1.5	0.3	1.9	0.3	37	12	76
MV07368 (F)	Mantoverde Sur	Rieger et al.. 2010	ICP-MS	8	9	369776	7061620	82.12	0.31	4.03	1.02	0.02	0.33	2.41	0.08	1.55	0.13	1.8	3.7	0.5	2.2	0.6	0.2	0.8	0.2	1.1	0.2	0.8	0.1	0.9	0.2	6	6	8
MV07577 (WR)	Mantoverde Sur	Rieger et al.. 2010	ICP-MS	6	10	369360	7060466	74.87	0.06	11.38	1.41	0.06	0.33	1.65	0.2	8.08	0.15	2	3.6	0.5	2.3	0.7	0.2	1	0.2	1.3	0.3	0.7	0.1	0.6	0.1	3	6	49
MV07176 (WR)	Manto Ruso	Rieger et al.. 2010	ICP-MS	5	4	369567	7065659	62.23	0.11	10.33	6.14	0.08	0.96	7.25	0.69	3.67	0.1	23	45.7	5.4	18.2	3.8	1	4.8	1	7.8	2	7.3	1.2	7	0.9	4	67	24
MV07206 (WR)	Manto Ruso	Rieger et al.. 2010	ICP-MS	4	11	369351	7065614	46.7	0.73	15.86	0.94	0.15	0.44	12.26	0.19	10.32	0.01	14.3	27.9	3.9	15.2	3.8	1.3	3.8	0.6	3.2	0.7	2	0.3	2.2	0.3	18	20	53
MV07135 (F)	Manto Ruso	Rieger et al.. 2010	ICP-MS	5	4	369567	7065659	76.09	0.29	12.46	1.67	0	0.29	0.06	1	5.73	0.01	0.6	1.4	0.2	0.7	0.2	0.05	0.2	0.1	0.4	0.1	0.4	0.1	0.7	0.2	4	4	20
MV07498 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	9	12	369769	7062449	51.65	0.81	17.64	8.35	0.33	5.44	7.59	4.11	1.78	0.16	13	28.5	4	16.3	3.9	1.2	3.3	0.5	3.1	0.6	1.7	0.3	1.6	0.3	28	16	466
MV07641 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	9	9	370085	7061965	50.1	1.12	14.48	8.15	0.16	5.98	5.41	0.72	6.64	0.72	19.9	46.7	6.2	24	5.8	1.4	5.9	1.1	7	1.3	3.7	0.6	3.4	0.5	37	34	64
MV07646 (WR)	Mantoverde Norte	Rieger et al.. 2010	ICP-MS	9	13	370784	7061713	55.63	1.03	18.77	2.94	0.11	3.98	10.65	5.2	0.57	0.02	5	9.9	1.5	7.1	2.1	0.9	2.2	0.4	2.8	0.6	1.6	0.3	1.6	0.3	33	16	407
MV07076 (WR)	East from Laura	Rieger et al.. 2010	ICP-MS	6	14	368711	7063983	50.39	0.96	14.43	15.98	0.24	5.43	0.5	0.13	5.99	0.29	4	7.1	0.9	4.1	1.1	0.3	1.5	0.2	1.5	0.3	1.2	0.2	1.3	0.2	13	10	92
MV07658 (WR)	Northwest from Celso	Rieger et al.. 2010	ICP-MS	6	12	369573	7066314	54.32	2.27	12.66	11.29	0.12	3.48	4.18	3.54	0.98	0.73	7.2	17.2	2.7	13.4	4.2	1.3	5.6	1.2	8.1	1.7	4.9	0.7	4.5	0.7	18	39	71
MV07058 (WR)	Quebrada Guamanga	Rieger et al.. 2010	ICP-MS	6	15	369756	7058510	85.79	0.26	4.74	2.14	0.14	0.28	1.73	0.1	1.71	0.05	0.6	1.6	0.2	1.1	0.3	0.1	0.6	0.1	0.7	0.2	0.6	0.1	0.8	0.2	3	4	12
AD0066-15	Alcaparrosa	Del Real et al.. 2018	LA-ICP-MS	10	16	374378	6962295	58.24	14.81	14.21	0.47	2.84	0.83	7.97	0.67	0.07	0.13	105	140	11.9	37.9	4.67	0.83	3.38	2.51	0.57	1.81	0.31	2.56	0.48	0.5	15.1	14.7	175
LD1493-18	Candelaria	Del Real et al.. 2018	LA-ICP-MS	10	16	373323	6956219	55.99	13.49	11.72	1.37	7.62	0.33	6.69	0.6	0.13	0.1	4.46	8.22	0.9	3.75	0.68	0.13	0.87	1	0.27	1.16	0.24	1.77	0.37	0.2	12.2	9.94	52.8
DH1077-10	Santos	Del Real et al.. 2018	LA-ICP-MS	11	17	376285	6961107	64.67	15.61	8.25	0.61	1.44	4.15	4.49	0.47	0.08	0.14	24.8	50.8	6.19	24.9	4.94	1.02	5.35	4.52	1.05	3.11	0.5	2.95	0.51	0.8	8.13	28.9	36.8
DH1077-12	Santos	Del Real et al.. 2018	LA-ICP-MS	11	17	376279	6961104	66.28	15.22	7.71	0.78	1.28	4.15	4.38	0.44	0.06	0.27	68.7	134	15.7	64.3	12.2	2.75	11.5	7.26	1.42	4.02	0.55	3.51	0.51	1.6	9.19	37.3	44.2
LD1493-2	Candelaria	Del Real et al.. 2018	LA-ICP-MS	12	18	373363	6956238	50.3	15.09	4.23	16.58	8.62	1.42	1.72	0.91	0.34	0.08	11.4	28.1	3.59	16.8	3.5	0.87	4.04	3.59	0.76	2.13	0.32	1.99	0.29	0.7	32.9	21.9	260
AD106-5	Alcaparrosa	Del Real et al.. 2018	LA-ICP-MS	12	18	374153	6962295	48.14	15.45	8.8	18.26	5.89	1.02	1.27	0.89	0.66	0.01	22.4	46.5	4.99	16.9	2.66	0.68	2.5	2.14	0.43	1.24	0.19	1.14	0.14	0.4	21.4	12.3	696
DH700-5	Santos	Del Real et al.. 2018	LA-ICP-MS	13	17	375752	6959326	65.61	14.35	7.04	0.62	1.56	1.67	6.39	0.58	0.07	0.17	1.62	5.78	1.12	7	2.45	0.55	2.81	1.38	0.69	2.32	0.38	2.64	0.51	0.5	13.2	19.9	214
ES064-GCI	District Exploration Drill hole	Del Real et al.. 2018	LA-ICP-MS	14	25	368960	6954611	70.07	15.07	3.77	0.56	0.55	6.56	3.29	0.28	0.01	0.09	4.84	12.9	1.69	6.65	1.44	0.24	1.31	1.65	0.39	1.24	0.2	1.48	0.26	0.2	4.03	11	52.6
b.1	Surface North	Del Real et al.. 2018	LA-ICP-MS	15	19	371338	6962255	58.57	17.23	7	5.33	2.6	3.93	2.92	0.75	0.14	0.27	23.7	51.7	6.62	28.9	6.33	1.38	4.95	4.34	0.86	2.6	0.37	2.61	0.42	0.8	15.1	24.6	461
S.2.8	Surface Alcaparrosa	Del Real et al.. 2018	LA-ICP-MS	15	19	372957	6961462	57.7	20.17	3.66	8.42	2.29	5.4	1.54	0.81	0.11	0.34	26.1	55.7	6.71	27.3	5.98	1.61	4.6	3.91	0.76	2.21	0.31	2.27	0.36	0.7	11.1	21.2	652
ES067 GC1	Exploration drill Hole	Del Real et al.. 2018	LA-ICP-MS	16	20	368896	6952881	61.65	16.93	7.45	2.55	2.49	5.52	3.3	0.47	0.07	0.14	18.3	36.7	4.49	18.4	3.48	0.92	2.96	2.52	0.54	1.67	0.24	1.63	0.3	0.5	10.3	15	188
ES067 GC2	Exploration Drill Hole	Del Real et al.. 2018	LA-ICP-MS	16	20	368954	6952880	59.54	17.82	6.74	5.81	2.47	4.65	1.8	0.5																			

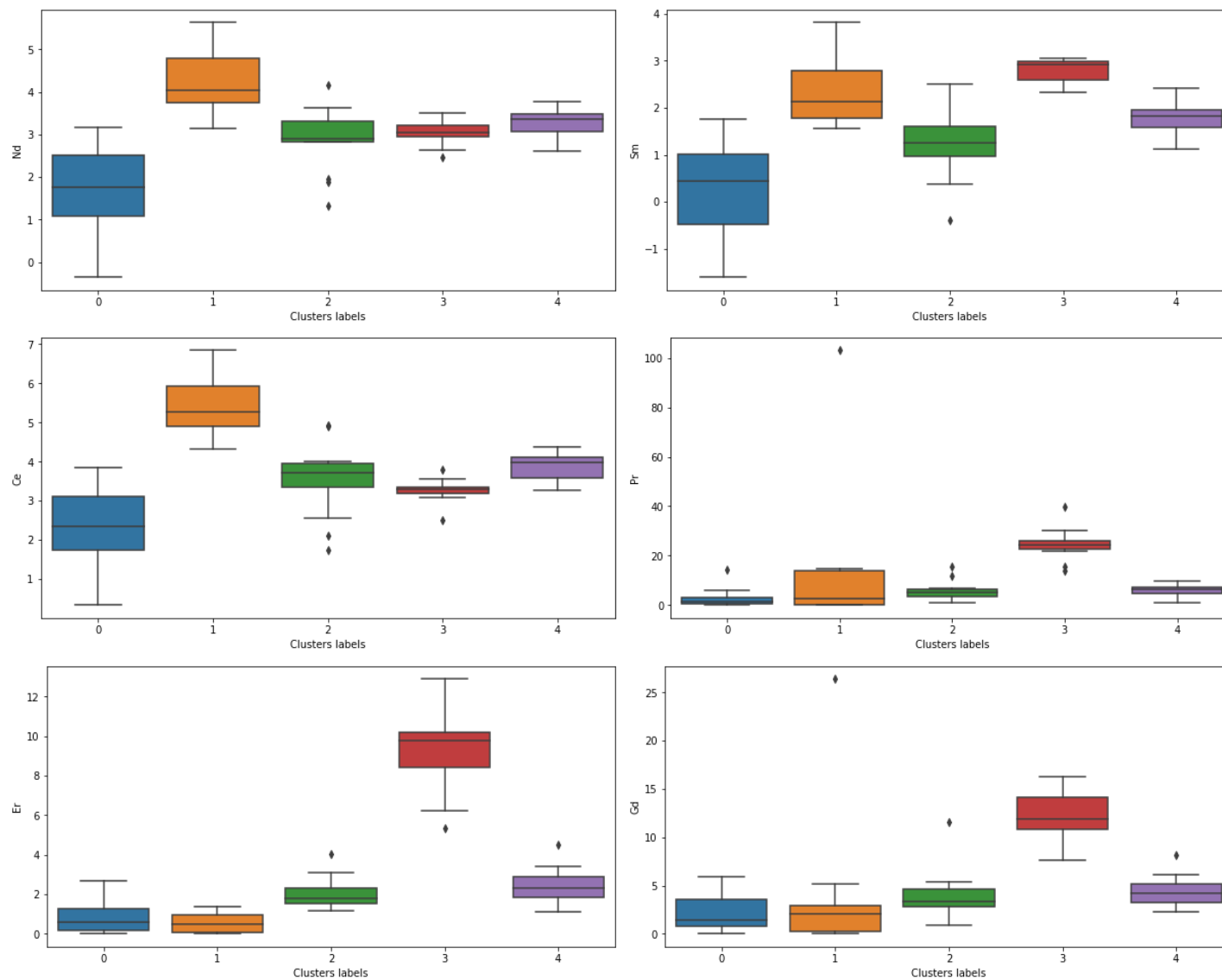
ANEXO 1.- (Continuación)

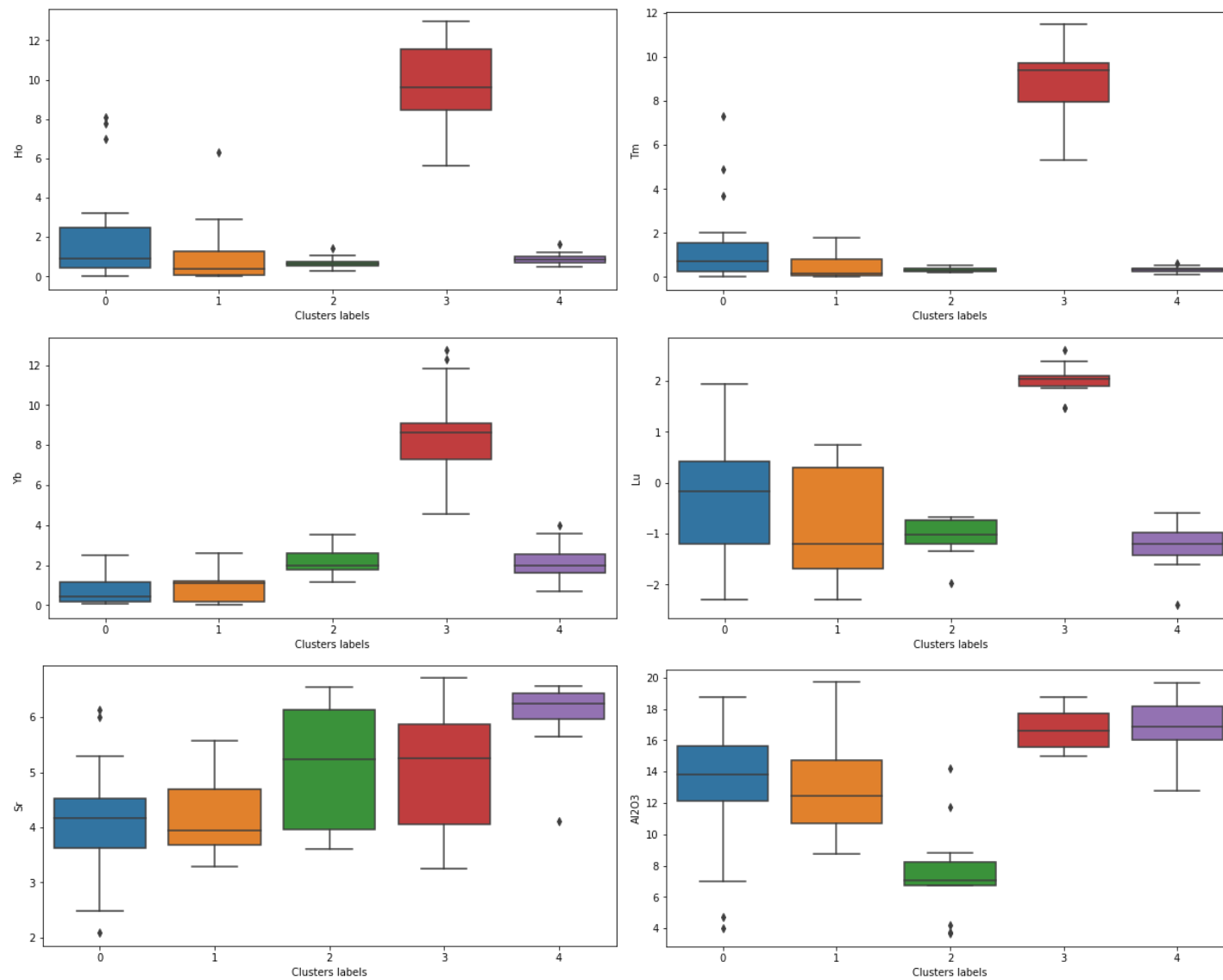
PC 862	Qda. Nantoco	Marschik y Fonboté 2001a	XRF ICP-AES	18	22	373190	6951620	52.43	1.13	18.74	8.53	0.21	2.97	8.13	3.32	1.17	0.18	27.7	24.5	24.4	22.7	18.7	17.4	15.6	12.8	10.9	9.78	9.71	8.64	7.35	NaN	NaN	22	820	
PC 904	Qda. Nantoco	Marschik y Fonboté 2001a	XRF ICP-AES	18	22	372000	6951370	53.71	1.02	17.69	7.85	0.17	5.1	8.66	2.92	0.79	0.14	29.9	26.1	25.2	21.1	18.7	14.3	15.2	12.8	11.7	9.78	9.4	9.09	8.24	NaN	NaN	20	190	
PC 1028	Qda. Las Pintadas	Marschik y Fonboté 2001a	XRF ICP-AES	18	22	367410	6948270	53.69	0.71	17.86	7.09	0.04	4.7	8.26	5.38	0.57	0.17	32	28.6	22.8	19.1	13.3	10.5	8.33	6.71	5.64	5.33	5.29	4.55	4.41	NaN	NaN	22	57	
PC 1031	Qda. Las Pintadas	Marschik y Fonboté 2001a	XRF ICP-AES	18	22	368215	6948700	57.18	0.74	15.77	7.02	0.09	5.05	5.25	6.86	0.38	0.14	36.6	28.7	24.4	20.3	13.3	17.7	12.3	11.7	9.87	8.89	9.41	8.18	8.24	NaN	NaN	26	37	
PC 671	Trinidad Claim	Marschik y Fonboté 2001a	XRF ICP-AES	11	17	376781	6959446	65.88	0.55	15.76	6.34	0.03	0.96	0.54	7.52	0.88	0.17	61.3	65.7	44.7	37.6	24.1	17.9	15.6	12.8	10.9	9.78	NaN	9.55	9.41	NaN	NaN	26	36	
PC 781	Trinidad Claim	Marschik y Fonboté 2001a	XRF ICP-AES	11	17	377060	6958960	65.47	0.47	14.99	6.69	0.1	0.91	1.09	7.36	0.79	0.18	53.1	44.9	39.8	33	20.7	13.3	13	12	11.5	11.1	11.2	11.8	10.9	NaN	NaN	21	26	
PC 835	Qda. Meléndez	Marschik y Fonboté 2001a	XRF ICP-AES	11	17	375980	6961040	66.42	0.47	15.28	6.45	0.03	0.69	0.41	6.5	2.96	0.17	15.6	28	15.5	11.8	10.3	7.92	9.42	14	13	12.9	11.5	12.7	10.9	NaN	NaN	9	93	
PC 1003	Ladrillos	Marschik y Fonboté 2001a	XRF ICP-AES	11	17	376360	6962370	64.28	0.48	15.55	6.75	0.02	0.63	0.77	6.25	3.88	0.14	41.2	29.4	22	18.6	11.3	7.14	7.61	8.45	8.72	10.7	10.3	12.3	13.5	NaN	NaN	9	58	
PC 692	Qda. Trinidad	Marschik y Fonboté 2001a	XRF ICP-AES	10	16	376672	6959188	59.47	0.58	13.29	12.68	0.11	4.23	0.25	0.78	4.92	0.12	6.4	3.82	NaN	3.49	NaN	1.69	NaN	2.92	3.08	NaN	NaN	4.09	5.29	NaN	NaN	8	107	
PC 799	Trinidad Mine	Marschik y Fonboté 2001a	XRF ICP-AES	10	16	376782	6959242	58.67	0.68	16.33	9.25	0.11	5.25	0.35	4.63	1.39	0.12	7.93	4.8	NaN	3.33	NaN	1.56	NaN	2.62	3.85	NaN	NaN	4.09	NaN	NaN	NaN	29	56	
PC 805	N of Carola Mine	Marschik y Fonboté 2001a	XRF ICP-AES	10	16	376125	6957865	55.01	0.65	15.29	10.43	0.16	4.8	0.34	0.12	6.82	0.17	127	79.4	57.7	34.4	16.8	7.66	6.52	2.92	2.82	3.11	NaN	NaN	NaN	NaN	NaN	12	107	
PC 920	S of Carola Mine	Marschik y Fonboté 2001a	XRF ICP-AES	10	16	375520	6955920	55.43	0.75	16.1	9.59	0.1	1.6	1.14	0.15	10.99	0.14	12.5	9.36	8.94	8.1	7.88	4.94	6.52	6.71	5.77	5.33	NaN	5.41	6.18	NaN	NaN	20	88	
PC 1102	Qda. Tirana	Marschik y Fonboté 2001a	XRF ICP-AES	10	16	375853	6960015	58.9	0.63	15.02	9.48	0.1	3.97	0.83	3.04	5.9	0.1	9.76	NaN	NaN	4.92	NaN	1.3	NaN	3.21	3.59	4.89	5.59	5.91	6.18	NaN	NaN	24	193	
PC 604	Qda. Paipote	Marschik y Fonboté 2001a	XRF ICP-AES	19	25	376750	6972600	63.32	0.52	15.28	5.93	0.07	2.23	1.38	8.43	0.72	0.1	45.1	35	30.1	25.9	20.7	11.8	12	9.04	8.72	8.44	7.94	8.64	7.65	NaN	NaN	39	144	
PC 605	Qda. Paipote	Marschik y Fonboté 2001a	XRF ICP-AES	19	25	376750	6972780	61.63	0.6	16.6	6.76	0.14	3.22	0.99	5.68	3.39	0.11	30.8	24.9	23.6	21	19	10.4	10.9	8.75	8.46	8	7.94	8.18	6.76	NaN	NaN	10	357	
PC 1096	Paipote	Marschik y Fonboté 2001a	XRF ICP-AES	19	25	375830	6966280	54.47	1.18	17.04	9.25	0.28	5.74	3.8	6.4	0.43	0.08	64.3	55.1	52	44.8	NaN	16.8	22.1	16.6	13.5	12.4	11.2	10.5	7.65	NaN	NaN	17	228	
PC 1220	Paipote	Marschik y Fonboté 2001a	XRF ICP-AES	19	25	374400	6965580	49.66	0.9	17.43	10.91	0.31	6.25	5.21	5.31	0.64	0.07	29.6	24.1	NaN	22.2	NaN	15.3	12.7	9.91	7.82	7.56	6.18	5.91	4.71	NaN	NaN	NaN	NaN	NaN
PC 1221	Paipote	Marschik y Fonboté 2001a	XRF ICP-AES	19	25	374420	6965480	52.34	0.8	15.88	8.44	0.36	7.58	5	5.63	0.55	0.04	19.8	18.2	NaN	19.2	NaN	11.7	13.8	12.2	10.4	9.78	8.82	8.18	6.47	NaN	NaN	NaN	NaN	NaN
PC 1336	Candelaria Trinidad	Marschik y Fonboté 2001b	XRF ICP-AES	20	23	376700	6959200	41.69	0.62	12	21.98	0.16	6.36	4.39	0.17	4.51	0.47	19.7	35.7	3.2	13.9	2.8	0.57	1.6	2.1	0.46	1.3	0.24	1.6	0.26	0.5	13	20	105	
PC 1345	Candelaria Algarrobos	Marschik y Fonboté 2001b	XRF ICP-AES	18	24	376200	6956500	45.02	0.52	11.62	30.1	0.07	2.14	1.01	0.09	7.64	0.1	78.1	160	13.8	46.8	5.8	1.02	3	2.8	0.61	1.4	0.25	1.6	0.24	0.5	11.4	8	162	
PC 188	Carola Meléndez	Marschik y Fonboté 2001b	XRF ICP-AES	11	24	377000	6956500	54.05	0.53	12.68	15.19	0.05	2.29	0.29	0.06	8.54	0.12	3.9	9.1	1.2	5	1.3	0.35	1.2	1.5	0.41	1.2	0.24	1.7	0.31	0.5	8	10	42	
PC 198	Carola Meléndez	Marschik y Fonboté 2001b	XRF ICP-AES	11	24	377000	6956500	56.41	0.51	13.73	11.36	0.07	2.58	1.13	3.1	4.42	0.16	2.1	5.3	1	3.8	1.2	0.28	1.4	1.7	0.46	1.6	0.27	1.9	0.34	0.5	8	12	85	
PC 219	Carola Meléndez	Marschik y Fonboté 2001b	XRF ICP-AES	11	24	377000	6956500	46.05	0.46	11.01	19.4	0.1	6.21	1.34	0	4.98	0.15	69.8	134	10.6	33.4	4.8	0.69	2.2	1.8	0.4	1.1	0.18	1.1	0.17	0.5	12.6	11	71	
PC 704	Socavon Rampa Meléndez	Marschik y Fonboté 2001b	XRF ICP-AES	11	26	377000	6958000	51.21	0.51	8.59	21.14	0.12	4.09	2.91	0	0.08	0.23	6.6	12.9	1.4	6.6	1.9	0.56	1.6	2	0.36	1.1	0.17	1.1	0.17	0.5	9.1	12	17	
PC 834	Socavon Rampa Meléndez	Marschik y Fonboté 2001b	XRF ICP-AES	11	24	377000	6958000	46.71	0.72	15.02	15.81	0.1	2.76	1.28	0.04	7.68	0.15	11	25.2	2.4	10.1	2.1	0.42	1.3	1.4	0.38	1.1	0.18	1.3	0.21	0.5	14.7	7	42	
PC 1356	Trinidad Meléndez	Marschik y Fonboté 2001b	XRF ICP-AES	11	24	376700	6959200	53.33	0.46	12.48	17.61	0.09	4.41	1.05	0	6.09	0.14	115	197	14.8	57.1	6.4	1.15	2.9	1.9	0.41	0.1	0.17	1.1	0.17	0.7	6	13	264	
PC 888	Qda. Meléndez	Marschik et al 1996	XRF	12	27	376762	6961535	44.51	0.77	14.82	4.91	0.74	3.29	14.49	2.39	1.24	0.09	5.9	11.3	1.7	7.4	2.1	0.79	2.5	2.5	0.47	1.3	0	1	0.14	NaN	NaN	20	195	
PC 1057	Qda. Los Toros	Marschik et al 1996	XRF	12	27	370180	6948320	46.83	0.98	17.06	5.43	0.43	2.58	10.27	5.72	1.49	0.15	10	23.3	3.2	14.8	3.3	1.06	3.7	3.6	0.64	1.4	0.19	1.1	0.15	NaN	NaN	19	152	
PC 1443	NaN	Marschik et al., 2003	XRF ICP-AES	21	28	370300	6958330	57.47	0.82	16.41	7.44	0.17	3.43	6.43	3.82	3.05	0.24	19.7	47.9	5.6	22.7	4.8	1.26	3.3	3.8	0.74	1.8	0.26	1.6	0.24	NaN	14	30	487	
PC 1498	San Gregorio Mine	Marschik et al., 2003	XRF ICP-AES	21	28	370000	6958250	57.42	0.75	17.9	5.86	0.13	2.36	6.66	3.96	2.62	0.3	22.7	54.2	6.9	27.7	6.3	1.45	4.2	4	0.82	2.2	0.31	1.9	0.28	NaN	11	13	644	
PD 97001	w. Alcaparrosa	Marschik et al., 2003	XRF ICP-AES	21	28	372536	6960819	53.83	1	18.48	8.27	0.15	2.84	7.68	3.96	2.37	0.55	24.8	59.7	7.3	34.2	7.5	1.82	5.3	5	0.96	2.6	0.36	2.1	0.32	NaN	14	29	689	
PD 97010	Viñita Azul	Marschik et al., 2003	XRF ICP-AES	21	28	367010	6962166	57.18	0.82	16.38	7.97	0.13	3.13	6.21	3.45	2.77	0.28	23.5	54.2	6.5	29.7	6.4	1.42	5	5	1.03	2.9	0.4	2.6	0.38	NaN	17	31	430	
PC 1307	NaN	Marschik et al., 2003	XRF ICP-AES	22	28	367600	6954900	56.37	1.06	17	8.08	0.11	3.34	7.47	5.07	1.21	0.56	25.7	58.6	7.9	33.2	8.3	1.73	6.1	5.2	1.03	2.8	0.36	2.3	0.35	NaN	22	35	478	
PC 1310	Q. Las Pintadas	Marschik et al., 2003	XRF ICP-AES	22	29	365420	6947980	59.9	0.52	18.23	3.69	0.14	3.18	8.43	4.24	1.05	0.17	15.5	34.5	4.3	21.4	5.2	1.24	4.3	4	0.74	1.9	0.26	1.6	0.24	NaN	16	26	692	
PC 97006	Candelaria-Panam.	Marschik et al., 2003	XRF ICP-AES	22	30	366700	6953860	55.78	0.97	18.23	7.33	0.08	2.38	6.68	4.61	2.28	0.46	29.9	67	7.5	36	8.1	1.97	6	5.9	1.23	3.1	0.43	2.7	0.41	NaN	16	35	570	
PC 97007	Q. Las Pintadas	Marschik et al., 2003	XRF ICP-AES	22	29	365360	6948220	55.36	0.56	18.79	7.65	0.12	2.93	7.08	4.51	1.16	0.22	14.4	31.4	4.3	17.8	4.6	1.33	3.1	3.3	0.71	2	0.3	1.7	0.25	NaN	16	21	711	
PC 97008	Q. Las Pintadas	Marschik et al., 2003	XRF ICP-AES	22	29	365372	6948218	56.95	0.52	18.09	7.17	0.1	2.87	7.08	4.29	1.02	0.2	13.8	30.5	4.1	17	4.4	1.21	3.3	3.3	0.7	1.8	0.28	1.8	0.26	NaN	16	21	663	
PC 1371	NaN	Marschik et al., 2003	XRF ICP-AES	23	25	364550	6953700	65.38	0.39	16.68	4.34	0.09	1.7	4.96	3.98	2.04	0.14	12.9	29.5	3.5	15.3	3.1	0.92	2.3	2.3	0.49	1.4	0.21	1.3	0.2	NaN	8	20	546	
PC 97009	Q. Los Lirios	Marschik et al., 2003	XRF ICP-AES	23	25	364520	6953320	62.96	0.41	17.04	5.03	0.09	1.81	5.33	4.13	1.74	0.16	12.4	26.5	1	13.6														

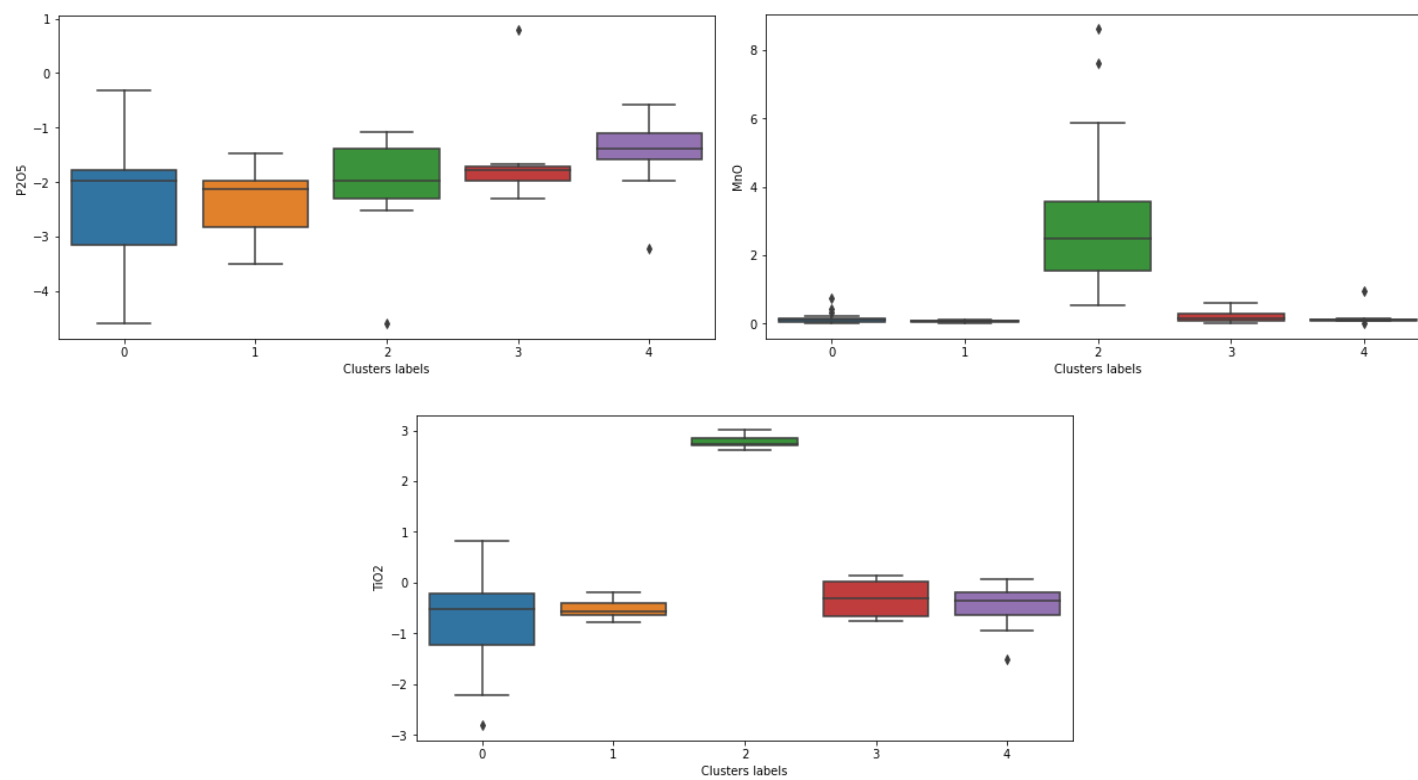
ANEXO 2.- Código

<https://www.kaggle.com/osciel/modelo-tesis>

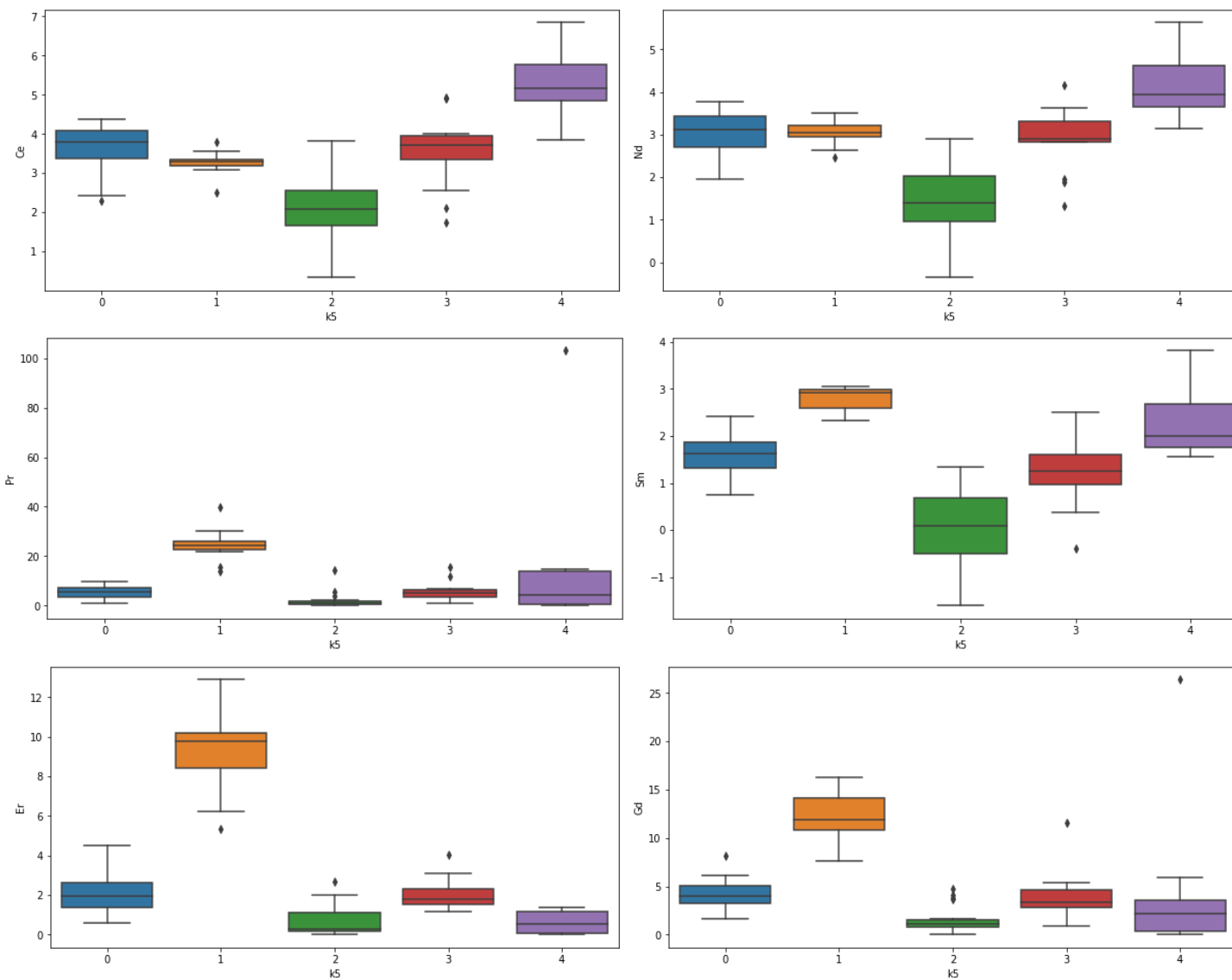
ANEXO 3.- Boxplots restantes de Agrupamiento Jerárquico.

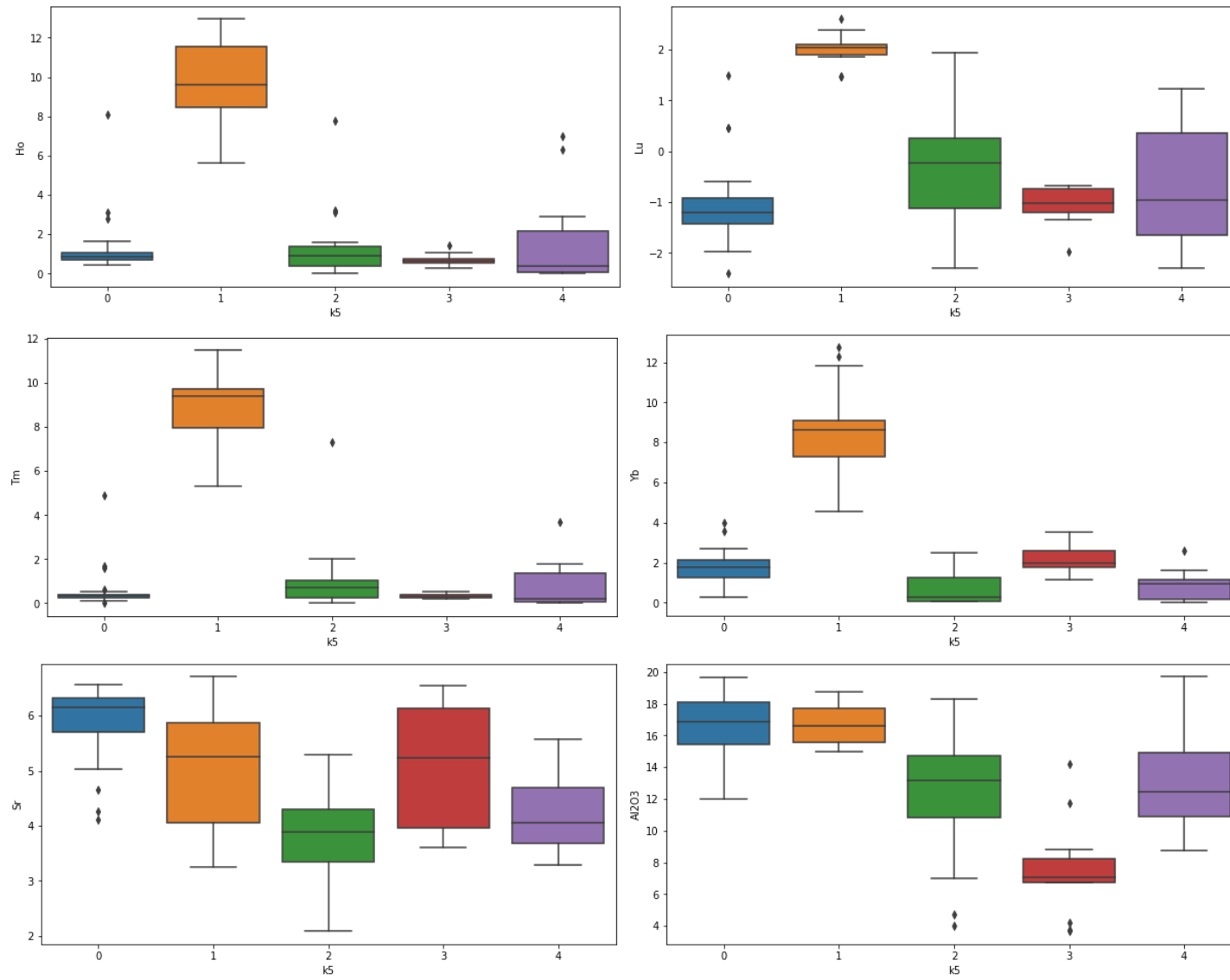


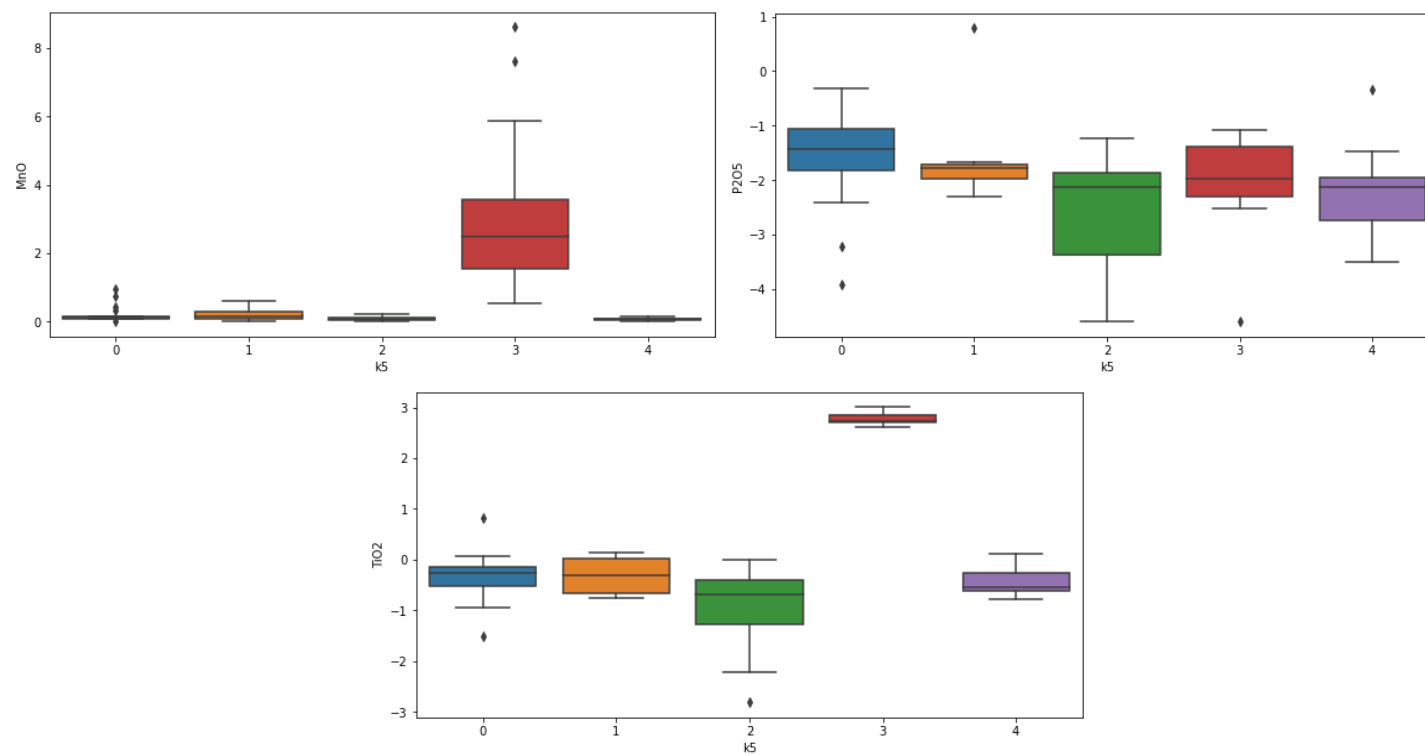




ANEXO 4.- Boxplots restantes de *K-means*.







ANEXO 5.- Tabla concentraciones promedio de REE en los clústeres con respecto a los valores de la corteza

Autor	La	Ce	Pr	Nd	Sm	Eu	Gd	Dy	Ho	Er	Tm	Yb	Lu
Wedephol (1995)	30	60	6,7	27	5,3	1,3	4	3,8	0,8	2,1	0,3	2	N/a
Lide (1997)	39	66,5	9,2	41,5	7,05	2	6,2	5,2	1,3	3,5	0,52	3,2	N/a
McGill (1997)	5 a 18	20 a 46	3,5 a 5,5	12 a 24	4,5 a 7	0,14 a 1,1	4,5 a 6,4	4,5 a 7,5	0,7 a 1,2	2,5 a 6,5	0,2 a 1	2,7 a 8	N/a
Clusteres													
AJC0	5.4	10.6	1.3	5.9	1.6	0.46	1.45	0.45	0.9	0.6	0.7	0.45	0.9
AJC1	115	196	2.7	57.1	8.5	1.8	2.1	0.6	0.4	0.5	0.18	1.1	0.3
AJC2	20.15	41.11	4.99	18.36	3.5	0.87	3.38	3.15	0.65	1.81	0.31	1.99	0.4
AJC3	30.49	26.94	24.39	21.11	18.72	13.77	11.96	11.66	9.62	9.78	9.4	8.64	7.7
AJC4	24.05	54.2	6.5	28.7	6.15	1.33	4.2	4.25	0.87	2.3	0.33	2	0.3
KMC0	19.7	44.9	5.6	22.5	5.05	1.25	3.95	3.9	0.87	1.95	0.33	1.75	0.3
KMC1	30.49	26.94	24.39	21.11	18.72	13.77	11.96	11.66	9.62	9.78	9.4	8.64	7.7
KMC2	3.9	8.1	0.9	4.1	1.1	0.3	1.1	0.2	0.9	0.3	0.7	0.3	0.8
KMC3	20.15	41.11	4.99	18.36	3.5	0.87	3.38	3.15	0.65	1.81	0.31	1.99	0.4
KMC4	99.45	178.4	4.45	51.95	7.45	1.65	2.15	0.85	0.405	0.55	0.215	0.95	0.4

ANEXO 6.- Boxplots para dos clústeres

